

TRUTH, REFERENCE AND REALISM

EDITED BY

ZSOLT NOVÁK
ANDRÁS SIMONYI



VISIT...

LANZAROTE
Caliente.COM

TRUTH, REFERENCE AND REALISM

TRUTH, REFERENCE AND REALISM

Edited by
Zsolt Novák
and
András Simonyi



Central European University Press
Budapest–New York

© 2011, the editors and the authors
T. Williamson: *Absolute Identity and Absolute Generality* © Timothy
Williamson, 2006
R. Wedgwood: *The Refutation of Expressivism* © Ralph Wedgwood, 2007

Published in 2011 by
Central European University Press

An imprint of the
Central European University Limited Company
Nádor utca 11, H-1051 Budapest, Hungary
Tel: +36-1-327-3138 or 327-3000
Fax: +36-1-327-3183
E-mail: ceupress@ceu.hu
Website: www.ceupress.com

400 West 59th Street, New York NY 10019, USA
Tel: +1-212-547-6932
Fax: +1-646-557-2416
E-mail: mgreenwald@sorosny.org

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted,
in any form or by any means, without the permission
of the Publisher.

ISBN 978-963-9776-86-9 cloth

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Truth, reference, and realism / edited by Zsolt Novák and András Simonyi.
p. cm.

Includes bibliographical references and index.

ISBN 978-9639776869 (hardbound)

1. Truth—Congresses. 2. Reference (Philosophy)—Congresses. 3. Realism—
Congresses. I. Novák, Zsolt. II. Simonyi, András.

BD171.O94 2005
121—dc22
2010027356

Printed in Hungary by
Pátria Nyomda, Budapest

Table of Contents

Acknowledgements	vii
Introduction	ix
<i>Zsolt Novák and András Simonyi</i>	
The Reality of Mathematics and the Case of Set Theory	1
<i>Daniel Isaacson</i>	
Conceptualism and Knowledge of Logic: A Budget of Problems	77
<i>Nenad Mišćević</i>	
What is Logic?	125
<i>Ian Rumfitt</i>	
Absolute Identity and Absolute Generality	177
<i>Timothy Williamson</i>	
The Refutation of Expressivism	207
<i>Ralph Wedgwood</i>	
Benacerraf's Problem, Abstract Objects and Intellect	235
<i>Howard Robinson</i>	
About the Authors	263
Index	265

Acknowledgements

This volume is a collection of six papers, the original versions of which were presented by the authors at the first Oxford-Budapest conference at Central European University in Budapest, 28–30 April 2005. By organising this conference, we wished to enhance scholarly interaction between the two philosophical communities after spending a research year in Oxford within the OSI/FCO Chevening Scholarship Scheme in the academic years 2002–2003 and 2003–2004, respectively.

We would like to express our gratitude to everyone who assisted in the organisation and successful management of the event, in particular to Krisztina Biber, our department coordinator at CEU, who was always at hand when we badly needed her help. Special thanks go to the seven main speakers, Ferenc Huoranszki, Daniel Isaacson, Nenad Mišćević, Howard Robinson, Ian Rumfitt, Ralph Wedgwood and Tim Williamson, for their personal encouragement, contribution and flexibility in the organisational period, and to their commentators, István Bodnár, Greg Bognár, Ferenc Csaba, Katalin Farkas, András Máté, Márta Újvári and Zsófia Zvolenszky, for their adept comments on the delivered talks. The conference could not have been brought into existence without the warm spiritual support and sponsorship of CEU's then president and rector, Yehuda Elkana. We are very grateful for this support as well.

The preparation of this volume was a longer and harder task than the organisation of the conference. First of all, we are grateful to CEU Press for declaring interest in the publication of this material. Also, we would like to thank the six authors for their patience and remarkably constructive attitude toward our written comments on their works. We benefited much from the stimulating conversations that accompanied this editorial work. Finally, we wish to express our gratitude to Linda

Kunos and Nóra Vörös, our editors at CEU Press, who provided invaluable help to us in preparing the typescript's final form.

We owe a different sort of debt to our families and friends, who are always with us when we occasionally leave our computers.

ZSOLT NOVÁK
ANDRÁS SIMONYI

Introduction

It is an entrenched and plausible view in philosophy that we can gain knowledge of objective truths by evidence other than sense experience. The clearest candidates of this type of knowledge are our claims about non-spatiotemporal domains, as in pure logic and mathematics, and those expressing analytic truths, independently of whether their intended subject matter is abstract or spatiotemporal. Beyond these paradigm cases, there are some other, more contestable examples as well, including our normative claims or value judgments in ethics, aesthetics and epistemology, and the descriptive claims of metaphysics.

Once we believe in the possibility of *a priori* knowledge acquisition, it becomes natural to ask ourselves: what is happening here, *how* do we learn what is objectively and necessarily the case without relying on the deliverances of sense perception? Clearly, any response to this question will draw heavily on what can be reasonably thought of the nature of those conditions whose obtaining or absence is supposed to determine the truth-value of the relevant claims. A proper explanation of *a priori* knowledge requires an appropriate conception of the meaning of *a priori* claims and the nature of *a priori* truths.

In philosophy of mathematics, the mutual dependence of theories of meaning and truth, on the one hand, and theories of knowledge acquisition, on the other, has long been an established part of common sense. This is in great part due to two seminal papers by Paul Benacerraf, published in 1965 and 1973, which have influenced virtually every writer on the subject since.

The first of these (Benacerraf 1965) became the groundbreaking work of mathematical structuralism. According to a structuralist, the

subject matter of mathematical theories is not a single domain of abstract individuals that, beyond having certain relations to each other, also possess some further properties, which distinguish the system in which they feature from other isomorphic ones. Rather, it is either all systems of individuals exemplifying an abstract structure or the structure itself whose elements are merely positions in the structure lacking any further individuating property, so that questions about what mathematical objects really are cannot be answered beyond what the theory says about the defining relations of these objects to each other.

The second paper (Benacerraf 1973) explicates a dilemma, which can be seen as fuelling many of the debates and inventions in early twentieth-century works on the foundations of mathematics. The dilemma is the following. If we maintain that the truth-value of our mathematical beliefs is determined by the obtaining or absence of those abstract and non-epistemic conditions that these beliefs purport to be about (i.e. whether certain mathematical objects possess certain mathematical properties), then we find ourselves unable to understand how we can, by means of our natural cognitive mechanisms, discover whether or not these conditions obtain. On the other hand, if we suppose that knowledge requires appropriate causal contact between knowing minds and the obtaining truth conditions of true beliefs, then we seem to be forced to conclude that the truth conditions of our established mathematical theories cannot be construed along the standard referentialist lines. Summing up, in philosophy of mathematics, our standard referentialist conception of truth seems to be incompatible with our standard causal theory of knowledge.

The concerns of the two articles are intimately related to each other. On the one hand, in a referentialist semantical framework our ideas of the subject matter of mathematics may be highly significant for our theory of mathematical truth. Some structuralists, for instance, believe that their conception of the subject matter of mathematics also resolves the explanatory puzzle about mathematical knowledge acquisition. On the other hand, an acceptable answer to Benacerraf's dilemma can provide us with reasons for taking a stand on mathematical structuralism as well. Some philosophers, for instance, believe that any acceptable answer to the dilemma must involve the rejection of semantical Platonism, i.e. the realist construal of the abstract subject matter of mathematics, which move may influence their views in the

debate over the structuralist understanding of mathematical objects and properties.

The significance of Benacerraf's observations is not confined to the philosophy of mathematics. Similar questions can be raised in the semantics of any discourse in which we are supposed to acquire knowledge of causally inert subject matters. Logic, for instance, is mostly supposed to concern inferential relations among propositions. Is there anything more to being a proposition than having certain inferential relations to other propositions? Are propositions, together with their inferential relations, real entities existing in an abstract (Platonic) realm, or are they merely projected by human minds? Can we maintain that the truth-value of our logical beliefs is determined by the obtaining or absence of those mind-independent conditions that these beliefs purport to be about? If we maintain this referentialist construal of logical truth, can we properly explain how our reasoning capacities could inform us about the obtaining or absence of such causally inert truth conditions? Isn't it the case that any causal account of how we actually *discover* objective logical truths undermines the adequacy of the standard referentialist construal of the truth conditions of logical beliefs?

The six papers collected in this volume address one or another of these questions in the context of mathematical (Isaacson), logical (Rumfitt, Mišćević) and ethical (Wedgwood) beliefs, and in the case of universal generalizations (Williamson) and beliefs involving ideas of universals in general (Robinson). In the remaining part of this introduction, we will briefly review the main theoretical options that one can adopt in response to Benacerraf's dilemma in the semantics of the above problematic discourses, and then provide a short overview of how the positions discussed in the six papers can be located on this theoretical landscape.

* * *

One way to start a review of the conceivable responses to Benacerraf's dilemma is to identify the crucial presuppositions of the case. Corresponding to the two horns of the dilemma, we can classify these assumptions into two major categories: semantical and epistemological assumptions.

The most fundamental semantical assumption behind Benacerraf's original case is that mathematical claims express genuine propositions that are truth-apt, some being true while others false (1973, 666). We shall call this first tenet *cognitivism* in the semantics of mathematics and the other problematic discourses in general.

The second semantical assumption underlying Benacerraf's dilemma is that metaphysical and epistemological considerations may impose substantive constraints upon a proper theory of meaning and truth. In particular, truth and falsity are substantive properties that play an important explanatory role, among others, in our account of knowledge acquisition about various domains (661, 662, 671). We shall call this second tenet *substantivism about truth* in the semantics of the relevant discourses.

Benacerraf's third semantical assumption is that truth in mathematics is a real, non-epistemic property (664, 665, 668, 674, 675, 676). In other terms, the truth conditions of mathematical claims obtain (or not) independently of anyone's actual knowledge of, or capacity to recognize, this particular circumstance, so no epistemic fact involving the truth-value of a mathematical claim is constitutive of the obtaining or absence of the claim's truth conditions. An ideal thinker can still be claimed to be able to know all mathematical truths, but the conceptual ground of this claim is not an epistemic construal of truth, but instead a realist construal of being an ideal thinker. Generalizing from the mathematical case, we shall call this third tenet *realism about truth* in the semantics of discourses about causally inert subject matters.

The fourth semantical assumption, explicitly discussed in Benacerraf's paper, is that the truth conditions of mathematical claims can be specified in terms of the intended subject matter of these claims (i.e. in terms of mathematical objects possessing mathematical properties) (665, 672, 677, 678). The assumption is independent of the previous two, since it does not imply anything substantive concerning the nature of the intended subject matters (664). What it does imply is adherence to the received referentialist construal of mathematical truth in conformity with our notion of truth in the semantics of other segments of natural language. Following Benacerraf's terminology, we shall call this tenet, generally, *referentialism about truth*, emphasizing that the term 'referentialism' has no substantive metaphysical implications here (i.e. that an advocate of this tenet need not commit herself

to any conception concerning the metaphysical status and nature of the relevant subject matters). Putting stress upon the conceptual independence of this tenet from the former two may be significant in the light of two relatively entrenched terminological conventions in present-day philosophy: first, the characterization of construals explaining meaning and truth without any reference to intended subject matters as anti-realist accounts in semantics; and second, the characterization of reference as a substantive relation between representations and represented entities, a notion clearly distinguishable from the deflated concept figuring in the label we suggest here for the received (broadly Tarskian) conception of truth.

The fifth semantical assumption, also explicitly touched upon by Benacerraf, is that the subject matters of mathematical expressions are the kinds of entities they are normally taken to be (673, 675). For instance, numbers and geometrical objects are abstract individuals that are causally inert and have no location in physical space and time. Again, this assumption is clearly independent from the earlier ones. One may maintain that mathematical claims are about abstract (i.e. non-spatiotemporal) states of affairs without subscribing to a substantive, realist interpretation of mathematical objects and properties and also without adopting the referentialist idea that the truth conditions of these claims have to be understood in terms of this abstract subject matter. Again, generalizing from the mathematical case, we shall call this fifth tenet *non-revisionism about subject matter* in the semantics of the relevant discourses.

On the epistemological side, Benacerraf's most fundamental assumption is that at least some of our mathematical beliefs qualify as knowledge (673). We shall call this first epistemological tenet, properly generalized, *anti-skepticism* in the epistemology of the relevant discourses. Since knowledge presupposes the truth of the known proposition, this assumption also implies that the truth conditions of at least some of our mathematical beliefs actually obtain.

The second epistemological assumption behind Benacerraf's case is that the acquisition of knowledge requires an appropriate causal link between the knowing mind and the obtaining truth conditions of the known propositions (671, 672). We shall call this second tenet *a causal theory of knowledge acquisition* in the epistemology of the relevant discourses.

Due to its negative reception in the subsequent literature, the previous assumption is today sometimes replaced by another, which holds that there must be some information-conveying contact between the obtaining truth conditions of known propositions and the knowing minds. The assumption rests on two fundamental convictions: first, that without invoking the existence of such a contact one cannot reasonably explain the reliability of beliefs; and second, that the impossibility of such an explanation undermines any epistemic ground one might have in support of those beliefs (Field 1989, 25–26). We shall call this weaker version of Benacerraf's second epistemological tenet *a contact theory of knowledge acquisition*.

Adopting the five semantical assumptions, we must conclude that the truth conditions of our claims about causally inert subject matters obtain (or not) without exerting any influence upon other existents, including our knowing minds. Adopting the two epistemological assumptions, on the other hand, we must conclude that at least in some cases there is an information-conveying mechanism between the obtaining truth conditions of our claims about causally inert entities and our actual evidence in support of these claims. The two conclusions clearly contradict each other: while the semantical assumptions suggest that there can be no contact between the truth conditions of claims about causally inert subject matters and the human minds, the epistemological assumptions imply that at least in some cases this contact obtains.

Conceivable responses to Benacerraf's dilemma can be classified also into two major categories: those that reject some of the semantical assumptions specified above, and those that abandon some of the epistemological assumptions.

Among the semantical responses, the most radical is the rejection of cognitivism concerning the problematic types of claims. If a claim is not an endorsement of a genuine proposition, thus it cannot be true or false, then it cannot qualify as a piece of genuine knowledge either. Of course, the systematic nature of our linguistic practice may still call for a proper explanation, but this account need not involve reference to the obtaining of any truth conditions. The best example of this *non-cognitivist* treatment of an otherwise problematic discourse is Hare's prescriptivism in metaethics (1952), but the same strategy has been traditionally attributed to metaethical emotivists, such as Ayer (1946) and Stevenson (1944), and more recently, by some authors, to

metaethical expressivists, like Blackburn (1993) and Gibbard (1990). In philosophy of mathematics, Hilbert's instrumentalist account of what he called ideal (infinite) mathematics (1925) is sometimes regarded as an instance of non-cognitivism in the sense specified above. On the current understanding, the crucial tenet of non-cognitivism is that the linguistic practice under scrutiny does not serve the expression of genuine beliefs, because it is not regulated by the detection of the obtaining or absence of some semantically significant conditions that could be regarded as conditions of truth.

A less radical semantical response to Benacerraf's problem is to deny the correctness of the second, substantivist assumption, and adopt a *deflationist* position in the semantics of the relevant discourses. Deflationists maintain that a proper theory of truth and reference is orthogonal to both our conceptions of the metaphysical status and nature of truth or correct declarative use conditions and our theories of how we acquire knowledge of the obtaining or absence of these conditions. In other terms, semantics is autonomous *vis-à-vis* metaphysics and epistemology. The main reason for this is that, on a deflationist understanding, truth is not a substantive property, so there is nothing to say about its nature and its relation to our epistemic capacities. Our notion of truth is fully characterized by the instances of Tarski's Disquotatation Schema or its counterpart for propositions as primary truth-bearers. A deflationist may still wonder how we can acquire knowledge of causally inert subject matters, and maybe even admit that, indeed, there is something theoretically puzzling in this phenomenon. Nevertheless, contrary to Benacerraf's claim, she can maintain that no response to this challenge can undermine the adequacy of referentialism about truth, since playing a substantive explanatory role in theories of knowledge is not a prerequisite for a condition to become constitutive of the truth conditions of a truth-apt representation. Classical versions of deflationism include Ramsey's redundancy theory (1927), Strawson's performative theory (1950), and Quine's disquotational theory (1970), while the most influential recent forms of deflationism are Grover, Camp and Belnap's prosentential (1975) and Horwich's minimal (1990) theories of truth. Beyond these clearly anti-substantivist examples, deflationist conclusions can be derived from Blackburn's (purportedly substantive anti-realist) "quasi-realist" program (1993) in semantics as well: if all distinctive claims of a realist can be endorsed,

on some re-interpretation, by an anti-realist as well, then it may seem quite natural to question the intelligibility of the very contrast between realism and anti-realism, and opt for a deflationist theory of truth.

The third available semantical reaction to Benacerraf's dilemma is to accept the cognitivist and substantivist assumptions, but deny the adequacy of realism, and adopt an *anti-realist* position about truth in the semantics of discourses about causally inert subject matters. Anti-realists about truth maintain that truth is a substantive epistemic property. In other terms, they hold that the truth conditions of a certain class of claims do not obtain independently of our capacities for recognizing these truths, or more exactly, that some epistemic facts concerning the truth-values of these claims are constitutive of the obtaining or absence of those truth conditions. Anti-realism in semantics and metaphysics has always found its basic motivation in epistemological considerations. No wonder that the doctrine may appear as a solution to Benacerraf's dilemma as well. If the truth-value of our claims about causally inert subject matters are construed in epistemic terms, then the explanation of knowledge acquisition need not invoke an information-conveying link between the knowing mind and something whose existence is fully external to it. Anti-realist replies may differ in their stance to Benacerraf's fourth and fifth semantical assumptions, i.e. whether they maintain or reject referentialism about truth and non-revisionism about subject matter in the semantics of the relevant discourses. It may be worth noting, however, that some influential doctrines from among those which are often classified as anti-realists about truth are arguably realist in the currently adopted sense of the term. Putnam's internal realist epistemisation of truth, for instance, is sometimes presented as a representative of (a referentialist and non-revisionist form of) anti-realism concerning this entity. Putnam, however, has never claimed that epistemic states are constitutive of the obtaining or absence of referential truth conditions.¹ Dummett's verificationist theory of truth cannot be regarded as anti-realist in the current sense either, since the conditions that he takes to be the truth conditions of our beliefs are supposed to obtain also independently of anyone's

¹ What the internal realist Putnam (1981) argues for is that the identity conditions of the intended states of affairs that can be regarded as the referential truth conditions of our beliefs are created by the classificatory work of mind.

actual knowledge of, or capacity to recognize, this particular circumstance.² Three further examples, whose anti-realist status is contestable, are Gibbard's projectivist semantics in metaethics (1990), Blackburn's quasi-realist construal of our claims about moral and modal states of affairs (1993), and Peacocke's conceptualism in the semantics of *a priori* discourses (2005). More plausible examples of anti-realism about truth include the construals of subjective idealists, Carnap's conventionalism about *a priori* (analytic) truth (1934), and maybe Brouwer's intuitionist theory in philosophy of mathematics (1949).

The fourth semantical response to Benacerraf's dilemma is to reject the referentialist assumption, and adopt *non-referentialism about truth* in the semantics of discourses about causally inert subject matters. Non-referentialists maintain that the truth conditions of a certain class of claims cannot be specified in terms of the intended subject matter of the constituents of these claims. For instance, on a non-referentialist construal, the truth conditions of mathematical claims are not mathematical states of affairs, whichever way these would be further understood. Rather, they are conditions that may or may not obtain in a non-mathematical realm. Non-referentialism does not imply anything about the metaphysical status and nature of the relevant subject matters. Nevertheless, it makes realism about truth compatible with anti-realism, fictionalism, eliminativism or quietism about subject matters. Of course, as Benacerraf rightly observed, an advocate of this position must explain what makes her preferred non-referential truth conditions qualify as conditions of *truth*. Once she can deliver this explanation, she can construe the relevant conditions, without reducing the corresponding subject matters, either in anti-realist or in epistemologically unproblematic realist terms. Examples of this non-referentialist strategy may include Dummett's verificationist construal of truth in discourses about epistemically inaccessible domains (1991), Blackburn's (1993) and Gibbard's (1990) projectivist theory in metaethics, and Putnam's (1967) and Hellman's (1989) modal structuralism in philosophy of mathematics. If conative attitudes, possession conditions of concepts and analytic links within our personal system of representation are

² What Dummett's (1991) anti-realist assumes is that the truth conditions of our beliefs are always verifiable (i.e. that we have an effective, though fallible, method to determine whether or not they actually obtain).

construed realistically, then a number of influential accounts that have been developed as an alternative to the epistemologically problematic realist and referentialist construals of truth can be classified as realist in the semantics of discourses about causally inert entities.

The fifth semantical strategy that may be adopted in response to Benacerraf's dilemma is to reject the fifth semantical assumption specified above, and embrace a *revisionist* construal of the referential domain of the problematic discourses under scrutiny. In the case of mathematics, for instance, this would amount to the view that mathematical claims are not about abstract states of affairs whose constituents are causally inert and have no spatiotemporal location, but instead they are either about some aspects of the natural world, or about some concepts in an active intellect, or about some other entities that can influence the human mind. Alternatively, a revisionist can take mathematical claims to be about *in re* or *ante rem* structures, rather than about a single system of individuals whose members, beyond having certain relations to each other, also possess some intrinsic properties that distinguish the system they constitute from isomorphic systems of other individuals. In other terms, she can take these claims to be either about all systems of individuals exemplifying a certain structure or about the structure itself that can be exemplified by those systems. Revisionism in itself does not imply anything about the metaphysical status of the relevant subject matters. A structuralist interpretation of mathematics, for instance, is compatible with a deflationist, an anti-realist and a realist construal of mathematical referents as well. Nonetheless, a major motive behind a revisionist construal of the subject matter of mathematics and other discourses about *prima facie* causally inert subject matters is that this construal allows for the wedding of a substantive realist and referentialist understanding of truth with a causal contact theory of knowledge acquisition in the philosophy of the relevant discourses. Theories falling into this class include Mill's and Kitcher's referentialist naturalism and various forms of structuralism in philosophy of mathematics (Mill 1843; Kitcher 1984). A short note about the structuralist proposals, however, may be in order. Although structuralist construals of Platonic entities are revisionist in character, not all of them support an effective response to Benacerraf's dilemma. In particular, those *ante rem* and *in re* structuralists who maintain that the problematic discourses under scrutiny are about causally inert and real

entities cannot claim that their view is more compatible with a causal theory of knowledge acquisition than the conception that a traditional (non-revisionist) Platonist provides.

In case one does not want to follow any of the five semantical strategies characterized so far, one may try to answer Benacerraf's dilemma by querying at least one of the epistemological assumptions of the case. The most radical epistemological strategy is to deny the possibility of knowledge about causally inert entities. If our received theories of such entities do not qualify as knowledge, then the adoption of Benacerraf's five semantical assumptions concerning the claims and implications of these theories remains compatible with the standard causal theory of knowledge. A limited form of skepticism may result from an error-theorist view of our received beliefs concerning causally inert entities. Such views have been defended by Mackie (1977) in metaethics and Field (1980) in philosophy of mathematics. An error-theorist argues that our received conceptions of a certain domain are false, and thus cannot qualify as knowledge, because the world does not contain those individuals and properties whose existence is required for their truth. Note, however, that an error-theorist need not assume that the existence of the relevant entities is a precondition of *any* truth about the corresponding domains. In absence of this assumption, she may maintain, for instance, that negative existential beliefs about causally inert entities are still true, and as such potentially qualifying as knowledge. A limited skepticism like this cannot resolve Benacerraf's dilemma. In order to save the compatibility of the standard realist and referentialist semantics with the standard causal theory of knowledge, one must deny the existence of any type of knowledge of the relevant causally inert domains.

The second epistemological strategy to follow in response to Benacerraf's case is to insist on the adequacy of the five semantical and the first epistemological assumptions, and query the idea that the acquisition of knowledge requires an appropriate causal link between knowing minds and obtaining truth conditions. Instead of admitting the adequacy of this causal account, the proponents of this position may argue that in the case of discourses about causally inert subject matters the contact between minds and obtaining truth conditions is not causal in character. We shall call this alternative a *non-causal contact theory of knowledge acquisition*. The classic example of this strategy is Gödel's

quasi-perceptionist account of mathematical knowledge (1944), while more recent instances of this category include Bonjour's account of rational insight (1998) and Brown's Gödelian view of mathematical knowledge and the nature of thought experiments (1991).

The third epistemological option one can choose in response to the dilemma is to deny the adequacy of the second epistemological tenet even in its weaker form, and subscribe to a *no-contact theory of knowledge acquisition*. Advocates of this position maintain that, although a proper explanation of knowledge acquisition may require an account of how the epistemic grounds of a knowing mind for adopting a certain class of true beliefs can reliably indicate the obtaining of the truth conditions of these representations, nevertheless, at least in the case of our discourses about causally inert subject matters, this account need not invoke the existence of any contact between these grounds and those conditions. Examples of this category include Wright's (1983) and Hale's (1987) neo-Fregean abstractionist, Balaguer's (1998) full-blooded Platonist, Katz's (1981) and Lewis's (1986) necessity-based, and Shapiro's (1997) and Resnik's (1997) structuralist strategy to account for the possibility of mathematical knowledge. In case one takes his holistic view of science together with his conception of ontological commitment seriously, Quine's empiricist epistemology (1948, 1951) also qualifies as a no-contact theory of mathematical knowledge.

Having reviewed the main theoretical options that can be adopted in response to Benacerraf's dilemma in the semantics of discourses in which we are supposed to acquire knowledge of causally inert subject matters, in the last part of this introduction, we shall provide a brief summary of the six papers appearing in this volume, and locate the positions they discuss on the theoretical landscape specified so far.



The first contribution, Daniel Isaacson's "The Reality of Mathematics and the Case of Set Theory" presents a structuralist account of mathematics that is realist about truth without being committed to the existence of mathematical objects.

Isaacson motivates his position by offering two reasons for rejecting the standard, Platonist semantics of mathematics in terms of particular abstract objects. The first is based on Benacerraf's dilemma:

Platonism is incompatible with the contact theory of knowledge acquisition. The second consideration, going back to Benacerraf's structuralist paper, is that for any allegedly correct interpretation of mathematical discourse in terms of abstract objects, there are other, isomorphic interpretations with equal claim to correctness, since mathematical truth is invariant under isomorphism.

According to Isaacson's alternative semantic proposal, both the subject matter of mathematics and the truth conditions of mathematical statements are constituted by particular mathematical structures, rather than mathematical objects, where the defining mark of particular structures is that all of their exemplars are isomorphic to each other. Particular structures (e.g. the structure of the natural numbers or that of the continuum) are contrasted with general structures (e.g. the structure of groups), which do have non-isomorphic exemplars.

Before expounding his own view on what particular structures are, Isaacson criticizes three rival accounts: the model-theoretic notion of structure, Stewart Shapiro's theory of structures, and modal structuralism. He points out that all of these approaches conceive of particular structures as constituted by objects (although of different kinds), thus contradicting the structuralist insight that the subject matter of mathematics consists exclusively of structures. He also argues that since model theory, and—at least on one interpretation—Shapiro's axiomatic theory of structures are themselves mathematical theories, it is circular, or the beginning of an infinite regress to take them as accounts of the subject matter of mathematical theories in general. As regards the last approach, modal structuralism, Isaacson is deeply skeptical about the role of modality in an *in re* account of particular structures: he thinks that there cannot be a difference between possible and actual existence for objects constituting the subject matter of mathematics.

Turning to his own positive account of structures, Isaacson articulates an *ante rem* realist view, according to which, instead of being constituted by objects, particular mathematical structures are given by coherent and categorical higher-order characterizations. When a characterization satisfying these conditions is established, the characterized particular structure also acquires real existence. In elucidating the notion of coherent characterization, Isaacson relies on Kreisel's conception of informal rigor: it is practicing mathematics through informal rigor which enables us to develop and understand descriptions of

mathematical structures and see their coherence. Claiming that establishing characterizations is an activity of the human mind, the paper arrives at the anti-Platonist thesis that mathematical characterizations are constituted by our concepts, and particular structures do not exist in advance of, or independently from these characterizations.

The semantics based on this conception of particular structures is tested on the foundationally central case of set theory. Isaacson argues that the second-order version of the Zermelo-Fraenkel theory of pure sets, which is quasi-categorical (its non-isomorphic models differ only in their “height”), coherently characterizes the structure of the cumulative set-theoretic hierarchy. Consequently, in his view, if a statement is semantically decided by this theory (in the sense that either the statement itself or its negation is a semantic consequence of the theory), then it has a determinate truth-value. Since Cantor’s continuum hypothesis is such a statement, Isaacson’s position implies that the continuum problem has a determinate answer, which set theorists may reasonably attempt to find.

Isaacson presents a revisionist and referentialist semantics, on which both the truth conditions and the subject matter of mathematical theories are constituted by particular structures. The proposed account is realist about mathematical truth and mathematical entities (although not about mathematical objects). The most important question left open by the paper concerns the metaphysical status of particular mathematical structures. On Isaacson’s account, particular structures are dependent on their conceptual characterizations. On the one hand, this dependence points towards a construal on which structures are constituted by our concepts. In this case, it is difficult to see how they could be different from their characterizations. (In fact, at one point Isaacson writes that structures *are* characterizations.) On the other hand, the fact that particular mathematical structures are considered to be *ante rem* types suggests a non-spatiotemporal interpretation, which would make the dependence on conceptual characterizations much more problematic. The two alternatives differ considerably from an epistemological point of view: Isaacson’s claim that we know mathematical structures through their characterizations and his commitment to the contact theory of knowledge acquisition can be easily combined with the view that structures are conceptual, but are difficult to maintain if structures are considered to be abstract (i.e. non-spatiotemporal) and real. If both the

epistemological advantages of the conceptual option and the intuition that characterizations and structures belong to different ontological categories are to be retained, then a non-referentialist construal of mathematical truth might be more congenial: such an approach could be realist about mathematical truth by claiming that the truth conditions of mathematical statements are constituted by characterizations adopted in space and time, while still holding that the subject matter of mathematics consists of non-spatiotemporal particular structures.

The second paper, Nenad Mišćević's "Conceptualism and Knowledge of Logic: A Budget of Problems" offers a criticism of conceptualism about logic, a view at the heart of recent attempts to answer Benacerraf's dilemma for logical discourse.

Adherents of conceptualism claim that logical knowledge is apriorist, and its source is our possession and mastery of logical concepts. Although conceptualists share these commitments, and many of them hold that the subject matter of logic is non-spatiotemporal, they still have a variety of theoretical options regarding the metaphysics of concepts, the role logical concepts play in the epistemology and semantics of logic, and the status of logical truth. From a metaphysical point of view, conceptualism is compatible with both realism and anti-realism about concepts, and also with a quietist rejection of both options. As regards the role of logical concepts, one can be conceptualist about justification, holding that the possession of logical concepts is a source of justification for logical knowledge, or about logical truth, characterizing the truth conditions of logical statements in terms of the properties and relations of our logical concepts (of course, the two options are compatible). Conceptualism about logical truth is a non-referentialist position, since the properties and relations of our logical concepts are not what the logical claims built from these concepts are about.

Since Mišćević adopts a broadly realist stance on logical truth, and takes the coincidence of truth conditions with subject matters for granted, the main targets of his criticism are philosophers defending a realist, referentialist, and justification conceptualist view of logic.

The paper examines three purportedly *a priori* sources of justification for logical laws that are prominent in the conceptualist literature. The first source is supplied by the influential thesis that respecting certain basic inference patterns is constitutive of possessing logical

concepts: many conceptualists maintain that this constitutive status can serve as the basis of an *a priori* justification of our belief in the validity of these standard inferences. In addition to this, conceptualist accounts of logical knowledge frequently rely on the obviousness and compellingness of basic inference patterns and principles, often in the form of claiming that finding these logical laws obvious and compelling is also a necessary condition for the possession of logical concepts. The third supposedly *a priori* source of justification is provided by the fact that logic is indispensable for any rational cognitive project.

Focusing mainly on Christopher Peacocke's version of conceptualism, Mišćević claims that the *a priori* justification envisaged by constitution-based accounts moves in an unacceptably narrow circle, and justifying our logical beliefs by a sufficiently wide equilibrium has to take into account the empirical success of our world-directed abilities to imagine certain complex situations while deeming others inconceivable. He also queries, following Timothy Williamson and Tyler Burge, the constitutivity assumption itself by pointing out that persons possessing logical concepts might fall short of the conceptualist requirements, e.g. because of having idiosyncratic views on logic.

Turning to the prospects of *a priori* justification through obviousness and compellingness, Mišćević argues that this characteristic is either a psychological-cognitive or a normatively characterized epistemological property. In the first case the connection between the subjective feeling of certainty and the purported objective warrant calls for an explanation, whereas in the latter an epistemological account of how we can reliably detect this normative property needs to be given. In both cases, Mišćević suggests, the required explanation would have an empirical character, making the amended justification *a posteriori*. In a similar vein, conceptualist accounts relying on indispensability are criticized on the ground that only cognitive projects with a good chance of success warrant their necessary preconditions, and the prospective success of a project is, to a large extent, an empirical matter.

Although his paper is mainly concerned with criticizing justification conceptualism about logic, Mišćević indicates that in his view the justification of logic is ultimately *a posteriori*, and is inherently related to the evolutionary success of our reasoning practices, with a special emphasis on the success of the "built in" ability to reason about concrete situations. From the perspective of Benacerraf's problem, the

main difficulty with this proposal concerns the ontological status of the truth conditions of general logical principles. On the one hand, if they are non-spatiotemporal, then how can we have *a posteriori* information about their obtaining? A well-known response to this problem, which Mišćević might be attracted to, is accepting a radical empiricist form of confirmation holism. This move, however, can hardly satisfy Benacerraf's curiosity, since an advocate of this doctrine would still owe us an account of how our empirical evidence could reliably indicate the obtaining of the purported abstract truth conditions of our logical beliefs. On the other hand, finding viable spatiotemporal truth conditions for principles of pure logic can be a daunting task in a referentialist framework, since our referential intentions appear to exclude the spatiotemporal construal of logical objects and properties.

As has been mentioned, Mišćević's arguments against conceptualism are targeted on approaches that are realist and referentialist about logical truth. This leaves open the question how effectively an aposteriorist could argue against accounts combining a concept-based, apriorist epistemology for logic with a non-referentialist, conceptualist construal of logical truth.

Ian Rumfitt's paper, "What is Logic?" develops a new, revisionist account of logic's subject matter, and applies it to the problem of explaining the epistemological role of deductive reasoning.

According to the standard view, logic investigates the unique relation of consequence that holds between the premises and the conclusion of a sound argument, and the *relata* of this relation are truthapt objects, which are, or can be, expressed by declarative utterances. Challenging this widely accepted picture, Rumfitt argues that we do not have a pre-theoretical grasp of a single relation of logical consequence, since the standards of soundness vary from context to context. Consequently, he claims, it is a mistake to describe logic as a science investigating a unique relationship. He also suggests that the existence of objects that are expressed by utterances should not be assumed early in an inquiry into the nature of logic, since our ordinary standards for assessing whether two utterances "say the same thing" are inadequate to sustain judgements of strict identity. In keeping with these reservations, Rumfitt adopts the traditional terminology and calls the *relata* of consequence relations propositions, but proposes to construe a

proposition as an ordered pair consisting of a declarative sentence type and a possible context of utterance.

Having criticized the standard view on logic's subject matter, Rumfitt turns to expounding his alternative account, according to which there exist a large number of different consequence relations, and logic is concerned with finding general truths valid for all of them. Consequence relations are set apart from other relations among the same type of *relata* by their modal characterization: for each space of possibilities there is a corresponding consequence relation that holds between a set of premises and a conclusion if and only if the conclusion in its actual sense is true in all possibilities of the space in which the premises in their actual sense are true. This modal conception of consequence relations differs not only from Russell's extensionalist construal of consequence as a relation that holds when either one of the premises is actually false or the conclusion is actually true, but also from Tarski's theory of consequence, which is formulated in terms of the actual truth-values of the premises and the conclusion under possible reinterpretations.

Rumfitt motivates his account by showing how it explains the epistemological role of deduction: if a thinker is reliable in deducing a conclusion from some premises only when the contextually relevant consequence relation holds, then she can apply her deductive capacity to premises she already knows, and splice together different pieces of knowledge in a conclusion whose truth is guaranteed by the consequence relation in question. In addition to being true, her belief in the conclusion will have been produced by a reliable belief forming procedure, satisfying one of the most important necessary conditions of knowledge. Rumfitt emphasizes that besides deducing conclusions from known premises, we can also attain new knowledge by reasoning deductively from premises that are *supposed* to be true. The conditions of soundness in these cases can be explicated only by invoking modal notions. Relying on this analysis of deductive capabilities and his construal of logical laws as general truths about consequence relations, Rumfitt accounts for the usefulness of logic by observing that learning a logical law enables us to extend our deductive capacity with respect to *all* consequence relations and gain new knowledge by deduction in every context we encounter.

Despite his rejection of the view that logic is concerned with a single consequence relation, Rumfitt holds that it is possible to identify

a broadly logical relation of consequence, which he characterizes, following Ian McFetridge, as being applicable and truth-preserving in the presence of any supposition. He also accepts, although with important qualifications, McFetridge's thesis that the correlative, broadly logical possibility is the weakest non-epistemic possibility: any proposition, which is possible in a non-epistemic sense (physically, metaphysically etc.), is possible in this weakest sense as well.

Rumfitt's referentialist account of logical truth characterizes the conditions of an argument's soundness and the truth conditions of logical laws in modal terms. At least some of these modal conditions are assumed to be real, as he emphasizes that broadly logical modality is not epistemic. Since non-actual possibilities are presumably causally isolated from the actual world, a variant of Benacerraf's dilemma might be applicable to Rumfitt's proposal: if the truth conditions of logical claims are causally isolated, then it is difficult to see how we could have a non-skeptical epistemology of logic that is compatible with the contact theorist account of knowledge acquisition.

Timothy Williamson's "Absolute Identity and Absolute Generality" argues that we can sharpen our understanding of certain issues concerning absolutely general quantification by drawing on analogies with corresponding issues concerning identity.

The paper starts with an interpretative question: what makes it the case that a speaker means identity by her predicate 'identical'? Williamson observes that if the speaker is committed to the reflexivity of 'identical,' and accepts the indiscernibility principle "if something is identical with something, then whatever applies to the former also applies to the latter," then—assuming that both of these commitments are correct, and her other words can be interpreted homophonically—it is provable that identity in her sense is coextensive with identity in our sense. Although the proof is elementary, it depends on the assumption that the quantifier 'whatever' in the speaker's indiscernibility principle ranges over a domain that includes identity in our sense. Williamson therefore turns to examining the problem in a more general setting, where the homophonic interpretation of words different from 'identical' is not taken for granted.

Formalizing the speaker's commitments in first-order logic, he points out that if the indiscernibility principle is construed as a schema

covering only formulas of the speaker's language, then there will be models that verify all commitments of the speaker, and still do not interpret 'identical' by identity. Williamson rejects excluding these non-standard models simply on the ground that they are not models for first-order logic with identity, and also criticizes Quine's methodology of interpretation, which requires us to interpret languages in a way that the strongest expressible indiscernibility relation is identity. Nonetheless, he argues that it would be a mistake to adopt Peter Geach's relativistic account of identity, according to which a predicate can play the role of identity only relative to a given language.

In Williamson's view, the key to formulating a satisfactory account of our grasp of absolute (i.e. not language-relative), identity lies in the fact that our commitment to the indiscernibility principle is open-ended: our predicate 'identical' refers to absolute identity at least partly because instead of treating the indiscernibility principle as limited to our current language, we have a general disposition to accept its instances in extensions of our language as well. In support of this account, Williamson recalls that reflexivity and the open-ended indiscernibility principle together uniquely characterize identity: if two predicates belonging to different first-order languages both satisfy these principles, then they are coextensive over the intersection of the two domains of quantification.

Turning to problems concerning absolute generality, Williamson establishes an analogous unique characterization result about universal quantification by proving that if two quantifiers belonging to different languages are both governed by open-ended analogues of the standard $\forall$ -introduction and $\forall$ -elimination rules, then they are logically equivalent. He claims that since the proof is purely syntactic, and does not depend on a semantic analysis of the universal quantifier, the result can be used to support the thesis that we have an idea of absolute, unrestricted generality in virtue of our commitment to the open-ended $\forall$ -introduction and $\forall$ -elimination rules.

Williamson formulates and answers a series of possible objections to this account of our grasp of absolute generality. Perhaps the most important among them is that the unique characterization argument presupposes the logical validity of the open-ended rules on the unrestricted reading, whereas they can be only materially valid, since the universal quantifier can be interpreted as ranging over any domain, not

exclusively over everything. In response to this objection, Williamson points out that we have to distinguish between the variable and the constant unrestricted accounts of the universal quantifier. According to the former view, for any things, the universal quantifier can be legitimately interpreted as ranging over just those things, while on the latter account the quantifier is mandatorily interpreted as unrestricted. Williamson's argument presupposes the correctness of the constant unrestricted account, as the open-ended rules are not logically valid on the variable construal.

Williamson does not argue against the variable account, but he notes that the analogy between absolute generality and absolute identity can help us to see that one argument against the constant unrestricted account is problematic. The fact that every interpretation that is legitimate on the constant account is also legitimate on the variable account but not *vice versa* cannot be used to support the claim that the variable account is preferable because of its greater generality, as the analogous argument for preferring first-order logic without identity to first-order logic with identity clearly fails.

The paper concludes with the observation that two frequently voiced objections to the absolute conception of generality are incompatible, as relativists cannot claim that generality absolutism is both inarticulate and inconsistent without undermining their own position. The incompatibility of the two charges is shown by an indirect argument: if absolutism was both inarticulate and inconsistent, then there would be a legitimate relativist interpretation of the universal quantifier on which everything accepted by the absolutist is true, and a valid proof of an explicit contradiction could be given from premises accepted by the absolutist. This proof would also demonstrate the inconsistency of generality relativism, as it would show that the relativist is committed to the existence of a legitimate interpretation on which an explicit contradiction should be true.

Although Williamson's paper makes no reference to this fact, there is an important link between absolutely general, unrestricted quantification and the structuralist approach to the semantics of set-theory. As we have already seen in connection with Isaacson's contribution, categorical theories play a key role in structuralist accounts, since they are considered as characterizing a unique structure, which determines the truth-value of every sentence of the theory's language. It has also been

mentioned that second-order Zermelo-Fraenkel set theory is only quasi-categorical. Nonetheless, an important result due to Vann McGee (1997) shows that the categoricity of set theory can be saved by switching from the standard variable construal of the universal quantifier to the constant unrestricted account: if only those interpretations are considered legitimate whose domain of quantification includes absolutely everything, then second-order Zermelo-Fraenkel set theory with urelements can be shown to be categorical in the sense that all of its models have isomorphic pure set hierarchies. In view of McGee's result, Williamson's vindication of the constant unrestricted construal of universal quantification can be seen as paving the way for a structuralist account of set theory that is capable of conferring a determinate truth-value on every set-theoretic statement.

Ralph Wedgwood's "The Refutation of Expressivism" argues against expressivist accounts in the semantics of normative discourse, and in favor of the rival, truth-conditional approach. Starting with a characterization of expressivism as the view that the fundamental explanation of the meaning of normative statements is to be given in terms of the mental states they express, without making reference to their truth conditions, the paper focuses on two variants of the position: emotivism and Allan Gibbard's semantic theory.

According to an emotivist, instead of representing properties, normative terms express certain feelings towards the evaluated objects or actions. Consequently, normative statements lack truth conditions and cannot be characterized as true or false. In response to this theory, Wedgwood recalls Peter Geach's critical observation that emotivism cannot account for the meaning of normative expressions in a variety of contexts, for instance when they occur within the scope of truth-functional sentential operators, or in propositional attitude attributions. Wedgwood argues that an adequate semantics has to explain how normative terms can occur in these contexts in the same sense as in others, must provide a uniform interpretation of the operators within whose scope they are embedded, and has to account for the fact that our customary inference patterns are valid in the case of normative statements as well.

Allan Gibbard's theory combines a quasi-realist stance on normative discourse (involving a thin interpretation of truth, proposi-

tions, properties etc.) with a psychologistic semantics on which normative statements express mental attitudes that can be characterized in terms of “hyperplans”: complete, consistent plans about what to do and think in every conceivable situation. Wedgwood admits that Gibbard’s theory satisfies the constraints imposed by Geach’s objection. Nevertheless, he claims that an adequate construal of our normative claims must also meet some further conditions. Building on Crispin Wright’s observations, he argues that our normative discourse is a disciplined enterprise, in which speakers aim to comply with certain standards of justification or warrantedness. Consequently, an acceptable semantics for normative claims has to provide an account of these standards as well. Wedgwood claims that any account satisfying this condition will assume that there is a property of normative statements, which is the point or purpose of normative discourse in the sense that speakers try to conform to the standards of justification and warrantedness because they aim at only accepting statements having this property. In addition, the property in question must have certain features that are possessed only by truth: among others, it is preserved in valid arguments, and satisfies an analogue of the T-schema. Wedgwood concludes that truth and truth conditions have an explanatory role in any adequate semantics of normative discourse, and consequently expressivism in general and in its specific Gibbardian form is untenable.

Wedgwood emphasizes that his discussion of expressivism relies on a substantive view about normative truth. In his book *The Nature of Normativity* (Wedgwood 2007), he develops a realist and referentialist truth-conditional semantics of normative discourse, and combines the revisionist thesis that non-physical normative facts can be causally efficacious with the view that in certain cases our knowledge of normative truths is based on *a priori* intuitions.

The motivation Wedgwood offers for classifying Gibbard’s account as expressivist shows that his notion of truth-conditional semantics is restricted to theories that explain meaning in terms of referential truth conditions. By contrast, his central argument against expressivism seems to establish only that an acceptable semantics of normative discourse has to refer to truth conditions *simpliciter*—he does not argue for the stronger claim that only referentialist construals can be adequate. As a consequence, Wedgwood’s argument does not rule out

the possibility of formulating an adequate truth-conditional, but nonetheless non-referentialist account that specifies the truth conditions of normative statements in the psychologistic terms characteristic of expressivist explanations of meaning.

The last paper of the volume, Howard Robinson's "Benacerraf's Problem, Abstract Objects and Intellect" argues that a variant of Benacerraf's dilemma can be raised with respect to any discourse, as all thought involves abstract objects in the form of universals. Robinson sketches a solution to the problem, according to which abstract objects exist only as concepts, and the ability of concept apprehension is a primitive and defining property of intellect.

The paper begins with a reconstruction of Benacerraf's original dilemma. On Robinson's interpretation, by pointing out that Platonism about mathematical truth and a naturalist, causal theory of knowledge acquisition entail that there is no mathematical knowledge, Benacerraf sets out a problem for philosophers who wish to combine realism about mathematics with naturalism. Robinson argues that it is possible to formulate a more general, discourse-independent variant of this challenge to naturalism: the ability to grasp abstract objects in the form of universals is a necessary precondition for thinking about any domain, but there is no viable, naturalistically acceptable account of this capacity. In support of this negative claim, Robinson criticizes the causal theory of conceptual representation, which he considers the standard contemporary naturalist solution to the problem of concept apprehension. In addition to pointing out that the causal theory ignores the connection between thinking and consciousness, the main problem he adduces is that an analysis of apprehension in terms of causal relations cannot account for the normativity of conceptual content.

Having argued against the feasibility of a reductive causal analysis, Robinson finishes his case against naturalistic theories of concept apprehension by dismissing the non-reductivist position that takes apprehension as primitive, but still maintains that grasping a universal necessarily involves causal contact with its instances. He claims that if we had a causally based primitive grasp of certain empirical features of the world, then *a priori* reflection on these features could acquaint us with a virtually unlimited number of further universals, but allowing

that the mind has an irreducible grasp of a vast range of (potentially uninstantiated) universals is incompatible with naturalism.

After his criticism of naturalistic accounts of concept apprehension, Robinson turns to examining the ontological status of universals. Surveying the three traditionally distinguished theoretical options, Platonism, *in re* realism, and conceptualism, he argues that neither of the realist approaches is viable, because Platonism cannot explain the predicative nature of the supposedly separate and self-subsistent universals, while *in re* realism is incompatible with realism about possibilities involving uninstantiated universals.

In line with his arguments against reductive theories of apprehension and realism, Robinson's own account of universals is a conceptualist view, according to which the ability to grasp concepts is a primitive and defining mark of intellect. Universals exist as concepts, but not as concepts of humans, because this would imply that universals unthought by humans do not exist, but as concepts of an objective intellect. Robinson indicates that this construal, which he terms a neo-Platonic account of universals, can be generalized to other classes of abstract entities, in particular to mathematical objects. On the general neo-Platonic view he advocates, the existence of abstract objects is independent from and presupposed by the existence of individual objects, and abstract entities exist as modes of understanding exercised by an objective intellect.

Relying on a revisionist conception of universals and abstract objects in general, Robinson provides a theory of *abstracta* that is realist from the human point of view, as no epistemic facts about humans are considered to be constitutive of the existence of abstract objects, but anti-realist from a global perspective, since abstract objects are supposed to exist only inasmuch as they have a role in rendering the world intelligible for an objective intellect. As a reply to Benacerraf's dilemma, the most important challenge Robinson's theory has to face is explaining how humans attain knowledge about abstract entities. In view of the strong intuitive appeal of a contact theory of knowledge acquisition, it seems questionable that by construing abstract objects as concepts of an objective intellect, neo-Platonists are significantly better placed to meet this challenge than proponents of traditional Platonism.

References

- Ayer, A.J. 1946. *Language, Truth and Logic*. 2nd rev. and reset ed. London: Victor Gollancz.
- Balaguer, M. 1998. *Platonism and Anti-Platonism in Mathematics*. New York: Oxford University Press.
- Benacerraf, P. 1965. "What Numbers Could not Be." *Philosophical Review* 74: 47–73. Repr. in Benacerraf and Putnam 1983, 272–294.
- . 1973. "Mathematical Truth." *Journal of Philosophy* 70: 661–679. Repr. in Benacerraf and Putnam 1983, 403–420.
- Benacerraf, P. and H. Putnam, eds. 1983. *Philosophy of Mathematics*. 2nd ed. Cambridge: Cambridge University Press.
- Blackburn, S. 1993. *Essays in Quasi-Realism*. New York: Oxford University Press.
- BonJour, L. 1998. *In Defence of Pure Reason: A Rationalist Account of A Priori Justification*. Cambridge: Cambridge University Press.
- Brown, J.R. 1991. *The Laboratory of the Mind. Thought Experiments in the Natural Sciences*. London: Routledge.
- Brouwer, L.E.J. 1949. "Consciousness, Philosophy, and Mathematics." In *Proceedings of the Tenth International Congress of Philosophy*, vol. 3, 1235–1249. Amsterdam: North Holland Publishing Company. Excerpts repr. in Benacerraf and Putnam 1983, 90–96.
- Carnap, R. 1934. *Logische Syntax der Sprache*. Vienna: Springer. Translated as *The Logical Syntax of Language*. London: Kegan Paul Trench, Trubner & Co., 1937.
- Dummett, M. 1991. *The Logical Basis of Metaphysics*. Cambridge, Mass.: Harvard University Press.
- Field, H. 1980. *Science without Numbers. A Defence of Nominalism*. Princeton, N.J.: Princeton University Press
- . 1989. *Realism, Mathematics, and Modality*. Oxford: Basil Blackwell.
- Gibbard, A. 1990. *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.
- Gödel, K. 1944. "Russell's Mathematical Logic." In *The Philosophy of Bertrand Russell*, edited by P.A. Schilpp, 125–153. Evanston: Northwestern University Press. Repr. in Benacerraf and Putnam 1983, 447–469.
- Grover, D., J. Camp, and N. Belnap, 1975. "A Prosentential Theory of Truth." *Philosophical Studies* 27: 73–125.
- Hale, B. 1987. *Abstract Objects*. Oxford: Basil Blackwell.
- Hare, R.M. 1952. *The Language of Morals*. Oxford: Clarendon Press.
- Hellman, G. 1989. *Mathematics without Numbers*. Oxford: Clarendon Press.
- Hilbert, D. 1925. "Über das Unendliche." *Mathematische Annalen* 95: 161–190. Translated as "On the Infinite," in van Heijenoort 1967, 367–392. Partially translated as "On the Infinite," in Benacerraf and Putnam 1983, 183–201.
- Horwich, P. 1990. *Truth*. Oxford: Basil Blackwell.

- Katz, J. 1981. *Language and Other Abstract Objects*. Totowa, N.J.: Rowman and Littlefield.
- Kitcher, P. 1984. *The Nature of Mathematical Knowledge*. New York: Oxford University Press.
- Lewis, D. 1986. *On the Plurality of Worlds*. Oxford: Basil Blackwell.
- Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong*. New York: Viking Press.
- McGee, V. 1997. "How We Learn Mathematical Language." *Philosophical Review* 106: 35–68.
- Mill, J.S. 1843. *A System of Logic*. London: Longmans, Green, and Company.
- Peacocke, C. 2005. "The *A Priori*." In *The Oxford Handbook of Contemporary Philosophy*, edited by F. Jackson and M. Smith, 739–763. New York: Oxford University Press.
- Putnam, H. 1967. "Mathematics without Foundations." *Journal of Philosophy* 64: 5–22. Repr. in Benacerraf and Putnam 1983, 295–311.
- . 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Quine, W.V. 1948. "On What There Is." *Review of Metaphysics* 2: 21–38. Repr. in Quine 1980, 1–19.
- . 1951. "Two Dogmas of Empiricism." *Philosophical Review* 60: 20–43. Repr. in Quine 1980, 20–46.
- . 1970. *Philosophy of Logic*. Englewood Cliffs, N.J.: Prentice-Hall.
- . 1980. *From a Logical Point of View*. 2nd ed. Cambridge, Mass.: Harvard University Press.
- Ramsey, F. 1927. "Facts and Propositions." *Proceedings of the Aristotelian Society, Supplementary Vol. 7*: 153–170.
- Resnik, M. 1997. *Mathematics as a Science of Patterns*. New York: Oxford University Press.
- Shapiro, S. 1997. *Philosophy of Mathematics: Structure and Ontology*. New York: Oxford University Press.
- Stevenson, C.L. 1944. *Ethics and Language*. New Haven: Yale University Press.
- Strawson, P. 1950. "Truth." *Proceedings of the Aristotelian Society, Supplementary Vol. 24*: 129–156.
- van Heijenoort, J., ed. 1967. *From Frege to Gödel: A Source Book in Mathematical Logic, 1897–1931*. Cambridge, Mass.: Harvard University Press.
- Wedgwood, R. 2007. *The Nature of Normativity*. Oxford: Clarendon Press.
- Wright, C. 1983. *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.

DANIEL ISAACSON

The Reality of Mathematics and the Case of Set Theory

1 The reality of mathematics

What is mathematics about? In what does the reality of mathematics consist? How can we know this reality? This paper propounds a realist conception of mathematics on which mathematical truth is objective but the truths of mathematics do not refer to mathematical objects. The subject matter of mathematics is structures (e.g. the structure of the natural numbers) rather than objects (e.g. the natural numbers). This conception is tested and illuminated by considering the case of set theory, both as a branch and as a foundation of mathematics.

There is an obvious answer to the first two of the questions with which we began (from metaphysics) that is so untenable when it comes to answering the third question (from epistemology) as to appear to refute the presumed reality of mathematics. This is that the objects with which mathematics concerns itself (e.g. natural numbers, real valued functions of a real variable, pure sets, points and lines of the Euclidean plane, etc.) exist, and the sentences about them are true in virtue of the properties of these objects, i.e. Platonism. If there are such particular objects, what possible contact can we have with them that would enable us to know something about them? Furthermore, the particularity of such supposed objects flies in the face of the evident fact that the truths of mathematics are invariant with respect to isomorphism. Rather, all three of our initial questions are to be answered by appeal to the notion of mathematical structure.

Mathematical structures are, roughly, of two kinds, particular (e.g. the natural numbers) and general (e.g. groups). Mathematics for its first several thousand years was concerned only with particular structures. Modern mathematics is much more about general structures,

but despite this shift, the reality of mathematics turns ultimately on the reality of particular structures. The reality of a particular structure, constituting the subject matter of a branch of mathematics such as number theory or real analysis, is given by its categorical characterization, i.e. principles which determine this structure to within isomorphism. A particular mathematical structure is not itself a mathematical object. The particular structures of mathematics constitute the determinate reality and objectivity of mathematics, and their role in understanding the nature of mathematics shows that the question whether mathematical objects exist is misguided. We have truth and realism without reference in mathematics.

This tension between the metaphysical attractions of Platonism and its epistemological intractability has been famously articulated by Paul Benacerraf in his paper “Mathematical Truth” (1973), and in current literature is often referred to as Benacerraf’s problem. Benacerraf had earlier claimed that considerations based on invariance under isomorphism show that there are no particular objects of mathematics: “numbers could not be objects at all; for there is no more reason to identify any individual number with any one particular object than with any other (not already known to be a number)” (1965, 69). In “Mathematical Intuition and Objectivity” (Isaacson 1994, 123) I endorsed this argument. However, as Jerrold Katz has pointed out, in “Skepticism about Numbers and Indeterminacy Arguments” (1996, 131–134) and in *Realistic Rationalism* (1998, 102–106), when advancing this argument one must not claim too much, namely that it establishes that there are no mathematical objects. Nonetheless these arguments do show that even if mathematical objects exist, their existence can play no role in answering our motivating questions. One might, by application of Ockham’s razor, conclude that they do not exist, but this conclusion is not required for the viewpoint I am advancing in this paper.

No categorical characterization of an infinite structure can be given in a first-order language. Categorical characterizations of infinite structures can be given in languages with second-order quantification, as Dedekind (1888) showed in the case of the structure of the natural numbers, the first such result to be established. Dedekind had published earlier, in 1872, a categorical characterization of the continuum, the first categorical characterization of any mathematical structure, but he did not prove its categoricity. In 1899 Hilbert gave a categorical

characterization of the Euclidean plane in his *Grundlagen der Geometrie*. In 1930 Zermelo established a corresponding result for set theory.

These proofs of categoricity are controversial since in their most natural formulations they use second-order quantifiers ranging over *all* subsets of the domain of first-order quantifiers, and there can be no complete logic for such second-order quantification.¹ Appeal to the notion of *all* subsets of the domain of first-order quantification is sometimes held to show that full second-order quantification depends on set theory, hence is not part of logic, and with the further difficulty when the characterized structure is itself the domain of sets that the characterization is circular. I shall argue that this claim is misguided in general and constitutes no particular problem in the case of set theory.

Establishing an account of set theory is a key test for any would be structuralist philosophy of mathematics. Set theory encompasses all of mathematics, in that all particular mathematical structures can be shown to exist within the cumulative hierarchy of pure sets.² It is in this sense that set theory constitutes a “foundation for mathematics.”³ Can set theory as a foundation of mathematics explain the reality of mathematics? Yes, if pure sets can be taken to be ontological atoms, since all the particular structures that constitute the reality of mathematics exist as equivalence classes under equivalence relations of isomorphism. But any such viewpoint ignores the fact that set theory is not only a foundation *for* mathematics but also a branch of mathemat-

¹ Categorical characterizations can be given using second-order quantification weaker than *full* second-order quantification, e.g. for arithmetic by using weak second-order quantification, where the second-order quantifiers range over all *finite* subsets of the domain of the first-order quantifiers. But equally there can be no complete finite logic of weak second-order logic. See Lopez-Escobar 1967, which presents “a formalization in which the proofs are of infinite length and which is complete”; see also Bell 1969.

² A possible exception might be category theory. Discussion of the relationship between category theory and set theory lies outside the scope of this paper.

³ Cf. Kreisel 1967, fn. 147: “But it is a *significant theorem* that the classical structures of mathematics occur already, up to isomorphism, in the cumulative hierarchy *without* individuals. For the reduction of mathematics to set theory it is important to convince oneself that intuitively significant features are invariant under isomorphism, or, at least, classes of isomorphisms definable in set-theoretic terms, e.g. recursive ones.” These results for the structure of the natural numbers and of the real numbers are included in most textbooks on set theory, e.g. Enderton 1977, chaps. 4 and 5; Levy 1979, chap. 2, § 4 and chap. 7; Jech 2003, chap. 4.

ics, and as such the status of set theory is in need of explanation as for any other branch of mathematics. Set theory as a foundation of mathematics is for these philosophical purposes no help, except to focus our attention on the case of set theory and to make it transparently clear that no set-theoretic understanding of structure can constitute a basis for a structuralist understanding of mathematics. In the first parts of this paper I expound a non-set theoretic understanding of the notion of a particular mathematical structure, and in the later parts I apply that understanding to articulate a structuralist understanding of set theory.

The key element of this understanding is Zermelo's categoricity theorems for second-order Zermelo-Fraenkel set theory.⁴ The problematic nature of the case of set theory for structuralism is immediate as soon as we consider how to formulate this result. It is natural to take categoricity of a theory as the property that any two models are isomorphic, where the notion of model is set theoretic, an understanding that is not available when the status of set theory is the issue. As we shall see, this problem can be addressed by exploiting a feature of the categoricity theorem for set theory which at first sight might seem a weakness but which is in fact intrinsic to the nature of the universe of sets, namely its unending extensibility, and in that way part of the strength of the result as fully characteristic of the fundamental notion of set, namely that the categoricity of set theory is not absolute: For any two models of ZFC², one of them is isomorphically embeddable in the other, but the embedding may be proper, i.e. one may constitute a larger universe of sets than the other, and the smaller will constitute a set within the larger domain. We shall also see that Shepherdson's (1951–1953) reformulation of Zermelo's results in terms of inner models of first-order von Neumann-Bernays-Gödel set theory clarifies the situation.

A byproduct of the categoricity for ZFC² is that despite independence from ZFC, Cantor's continuum problem has a determinate answer which requires new axioms to find, analogously to the need to extend Zermelo's axiomatization of set theory (1908) by the Axiom of Replacement to establish that every Borel set of reals satisfies the

⁴ I shall use a superscript 2 to signify the second-order formulation of an axiom system. Systems designated without a superscript 2 are first-order, e.g. ZFC² vs ZFC.

infinite game-theoretic property of determinacy (see Martin 1975 and Friedman 1971), or the need to extend Peano Arithmetic by transfinite induction of order-type ϵ_0 in order to prove Goodstein's Theorem (see Goodstein 1944, and Kirby and Paris 1982).

More recent results include categoricity of the theory of p -adic reals, and of the continuum with infinitesimals (this last result is not quite unequivocal: the characterization of the continuum with infinitesimals is provably unique to within isomorphism on the assumption that the continuum hypothesis holds; see Prestel 1990, 326).

2 The reality of constructive mathematics

As remarked above, I do not accept the revisionist claims of constructive mathematics, in particular intuitionism. More precisely, I accept constructive mathematics (of course) but not constructive philosophy of mathematics. I shall briefly sketch how the ideas of the previous section can be deployed to account for the reality of intuitionistic mathematics.⁵

The reality of intuitionistic mathematics is suggested by its historical development. The development of the intuitionistic theory of the continuum between 1907 and 1980 is comparable to the development of the classical continuum from 1660 to 1900. How to account for this objectivity? Consider the intuitionistic theorem that every continuous function from the reals to the reals is continuous. This truth is accounted for in terms of the (classical) structure of the continuum by seeing the intuitionistic theory of real valued functions of a real variable as a restriction of the classical theory, namely as those functions for which the rational approximation of the value of the function is determinable (computable) from the rational approximation of the input. This understanding is made mathematically precise by the condition of continuity.

⁵ These remarks amplify my discussion (Isaacson 1994, §5), which concluded with the declaration that intuitionistic mathematics is real and objective in the same way in which all mathematics is (135).

Similar considerations apply e.g. to strong intuitionistic refutations of excluded middle, i.e. interpretations in which a statement of the form

$$\neg \forall \alpha (\exists n \alpha(n)=0 \vee \neg \exists n \alpha(n)=0)$$

is true when α ranges over absolutely lawless sequences of 0 and 1. Being absolutely lawless, whatever is true of such an α is true on the basis of an initial segment of α , so if $(\exists n \alpha(n)=0 \vee \neg \exists n \alpha(n)=0)$, then there must be a natural number k such that $\exists n \alpha(n)=0$ or $\neg \exists n \alpha(n)=0$ is determined by the first k values of α . But there is no natural number k such that if no 0 turns up in the first k elements of a random sequence of 0 and 1 then a 1 will never turn up. This is as true of the classical understanding of infinite sequences of 0 and 1 as it is on the intuitionistic understanding. There is no tension here with the fact that $(\exists n \alpha(n)=0 \vee \neg \exists n \alpha(n)=0)$ is true on the classical meaning of $\vee$. If $A \vee B$ is interpreted as “ A is determined as true or B is determined as true, on the given information,” then for α ranging over lawless sequences, $\forall \alpha (\exists n \alpha(n)=0 \vee \neg \exists n \alpha(n)=0)$ is refutable, i.e. $\neg \forall \alpha (\exists n \alpha(n)=0 \vee \neg \exists n \alpha(n)=0)$ is true.

Considerations of this kind apply more generally to those parts of mathematics that are intensional in character, such as the theory of computation.

3 Mathematics as theories of structures

Understanding mathematics in structural terms began in the second half of the 19th century with Richard Dedekind’s determination, in 1858, to explain to his students at the ETH in Zurich the fundamental properties of the continuum (opening paragraph to Dedekind 1872). Dedekind identified the “essence of continuity” in the principle that “If all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes,” i.e. the cut property, and remarks that “if we knew for certain that space were discontinuous there would be nothing to prevent us, in case we so desired, from filling up its gaps in thought and thus making it continuous ... in accordance with the

above principle” (1872, 771–772). So far as I am aware no one earlier than Dedekind can be construed as understanding mathematics in structuralist terms.

Dedekind gave an explicitly structuralist account of the natural numbers in 1888 which shared fundamental insights, independently arrived at, with Frege’s *Grundlagen der Arithmetik*, most centrally Dedekind’s notion of chain and Frege’s notion of following in the ϕ -series. There were also fundamental philosophical differences. Frege was convinced of the particularity of the individual numbers, and considered it a scandal that mathematicians were unclear as to what they are. He was also convinced that the truths of arithmetic are analytic. These two views together led him to his notion of extensions of concepts and the fatal Basic Law V. In 1899, while Frege was at the height of his powers, bringing his program to completion and before the disaster of Russell’s paradox had been made manifest, Hilbert published his categorical characterization of the Euclidean plane (1899). Hilbert’s approach was avowedly structuralist. Frege was appalled and wrote to Hilbert to let him know. Hilbert’s structuralism was proclaimed in his slogan, “consistency implies existence.” Frege held the opposite view, that it is only in virtue of the existence of the numbers that arithmetic is consistent. Their correspondence brings these issues clearly into focus. (The English translations in the following quotations are from Kluge 1971.)

The correspondence began with a long letter from Frege to Hilbert on 27 December 1899 in which he declares:

it can never be the purpose of axioms and theorems to establish the reference of a sign or word occurring in them; rather, this reference must already be established. (Frege 1976, 62–63; Kluge 1971, 8)

From the fact that axioms are true, it follows of itself that they do not contradict one another. (Frege 1976, 63; Kluge 1971, 9)

Hilbert replied immediately, 29 December 1899, and responded to the line just quoted as follows:

You write “... From the fact that axioms are true it follows that they do not contradict one another.” I was extremely interested

to read just this proposition in your letter, because for as long as I have been thinking, writing, and lecturing about such things, I have always been saying the opposite: If the arbitrarily posited axioms together with all their consequences do not contradict one another, then they are true and the things defined by these axioms exists. For me, this is the criterion of truth and existence. (Frege 1976, 66; Kluge 1971, 12)

You say that my concepts, e.g. “point,” “between” are not unequivocally fixed; that on p. 20, for example, “between” is taken in different senses and that there a point is a pair of numbers.—But surely it is self-evident that every theory is merely a framework or schema of concepts together with their necessary relations to one another, and that the basic elements can be construed as one pleases. If I think of my points as some system or other of things, e.g. the system of love, of law, or of chimney sweeps ... and then conceive of all my axioms as relations between these things, then my theorems, e.g. the Pythagorean one, will hold of these things as well. In other words, each and every theory can always be applied to infinitely many systems of basic elements. (Frege 1976, 67; Kluge 1971, 13–14)

Frege in his long reply to Hilbert, 6 January 1900, amplified the point with which Hilbert had taken exception by declaring:

What means do we have for proving that certain properties or requirements (or however else one wants to put it) do not contradict one another? The only way I know of is to present an object that has all of these properties, to exhibit a case where all these requirements are fulfilled. Surely it is impossible to prove consistency in any other way. (Frege 1976, 70–71; Kluge 1971, 15)

And Frege challenged Hilbert with the Julius Caesar (or in this case his pocket watch) problem:

I do not know how, given your definitions, I could decide the question of whether my pocket watch is a point. (Frege 1976, 73; Kluge 1971, 18)

Hilbert politely declined to pursue the matter, pleading pressure of work in his reply of 15 January 1900.

Hilbert's ideas for purely syntactic consistency proofs (1904 and 1920s) call into question Frege's conviction that there is no other way to prove consistency of a mathematical theory than by establishing the existence of its objects, though not straightforwardly in light of Gödel's Second Incompleteness Theorem. Regardless, it may seem almost a truism that Frege is at least right that existence establishes consistency, given what now are called model-theoretic consistency proofs. Frege parlays this seeming truism into an argument against Hilbert that if the objects of a given theory do exist it must be redundant to establish their existence by proving the consistency of that theory. "However, if one has such an object, one would not need to prove that there is one by the roundabout way of proving consistency" (Frege 1976, 75; Kluge 1971, 20).

Frege is wrong in terms of the philosophy of mathematics at issue here. It is not strictly possible to prove consistency from existence, despite the seemingly obvious fact that we are convinced that e.g. Peano Arithmetic is consistent because it is true in the natural numbers. The point is that we cannot be convinced as to the existence of the natural numbers unless we are convinced that we have a coherent account of them. The coherence of any such account comes down to consistency of a mathematical theory. This is the essence of a structuralist philosophy of mathematics. Structuralism is a widely (even though by no means universally) accepted way to think about mathematics but so far as I am aware none of its proponents have grasped the (in these terms) radical point that *there are no mathematical objects*, only mathematical structures, which themselves are not objects. I shall develop this point later.

The two leading mathematicians at the end of the 19th century and beginning of the 20th were David Hilbert and Henri Poincaré. Poincaré, like Hilbert, held a structuralist view of mathematics. In 1902, Poincaré declared "Mathematicians do not study objects, but the relations between objects; to them it is a matter of indifference if these objects are replaced by others, provided that the relations do not change. Matter does not engage their attention, they are interested by form alone" (1902, 20). Like Hilbert, Poincaré was philosophically acute about mathematics, but unlike Hilbert did not develop his conceptions

of mathematics into mathematical programs, and indeed was anti-foundational and specifically critical of Hilbert's work on foundations.

Paul Bernays was Hilbert's chief collaborator in logic and foundations of mathematics from 1917, when he came to Göttingen from Zurich to become Hilbert's Assistant, until 1934, when the Nazis began to expel Jews from the universities and he returned to Zurich. He played a key role in articulating Hilbert's program, which received its definitive formulation in Hilbert and Bernays *Grundlagen der Mathematik*. Bernays wrote volume 1 (1934) with some input from Hilbert. Volume 2 (1939) is entirely written by Bernays. Bernays, like Hilbert, espoused a structuralist understanding of mathematics:

If we examine what is meant by the mathematical character of a deliberation, it becomes apparent that the distinctive feature lies in a certain kind of abstraction that is involved. This abstraction, which may be called formal or mathematical abstraction, consists in emphasizing and taking exclusively into account the structural aspects of an object, that is, the manner of its composition from parts; 'object' is understood here in its widest sense. One can, accordingly, define mathematical knowledge as that which rests on the structural consideration of objects. (Bernays 1930, 23, English trans. 7)

Bernays summarized this perspective with a statement that

our characterization of mathematics as a theory of structures seems to be an appropriate extension of the view mentioned at the beginning of this essay that numbers constitute the real object of mathematics. (1930, 32, English trans. 15)

In 1930, the same year in which Bernays published the paper from which I have just quoted, Zermelo published his paper characterizing the cumulative hierarchy of sets (Zermelo 1930). The Löwenheim-Skolem theorem had established that no first-order axiomatization of set theory could characterize the intended structure of sets, with its uncountable infinities, despite which Skolem and others were adamant that mathematics should be expressed and axiomatized in first-order languages. Zermelo's notion of "definite property" in the

Axiom of Separation in his 1908 axiomatization of set theory was to be construed as “property expressible in the first-order language of set theory.” Correspondingly, the Axiom of Replacement suggested by Fraenkel and Skolem, and implicit in a prescient letter from Cantor to Dedekind in 1899 was to be formulated as an axiom schema with an instance for each mapping from sets to sets definable in the first-order language of set theory. Zermelo did not accept the restriction of first-order languages, and seemingly never understood the considerations which motivated other logicians in the '20s and '30s to make this restriction. By rejecting this restriction and taking the Axiom of Replacement as a second-order axiom rather than as schema of first-order axioms, he was able to establish his categoricity result.

When Cantor conjectured that $2^{\aleph_0} = \aleph_1$, Cantor's Continuum Hypothesis (1878) immediately became the leading unsolved problem in set theory. The Continuum Hypothesis is equivalent to the claim that every uncountable subset of the reals is equinumerous with the reals. Cantor made some seeming progress on it by proving that all closed and open uncountable sets have power of the continuum. In his address to the World Congress of Mathematicians in Paris in 1900 Hilbert listed the continuum problem first among the twenty-three most important open problems of that time. Zermelo (1930) made no reference to the continuum problem, but in a report to the *Notgemeinschaft der Deutschen Wissenschaft* (published posthumously, in 1980, by Gregory Moore as an appendix to “Beyond First-order Logic”).⁶ Zermelo draws the inference that the continuum problem is shown to be determinate by his results in “Über Grenzzahlen und Mengenbereiche” (Zermelo 1930, see Moore 1980, 132), though his observation of this inference is marred by the fact that he claims too much, namely that the *generalized* continuum hypothesis is determined: After a statement of his isomorphism theorems from “Über Grenzzahlen und Mengenbereiche,” he says (in my translation): “It follows already, among other things, that the (generalized) Cantor conjecture (by which the power set of every set has the next higher cardinality) does *not* depend on the choice of models, but through our axiom system is always decided (as being true or false).”

⁶ Moore dates this report as “sometime between 1930 and 1933, and probably at the beginning of that period” (Moore 1980, 124).

In 1938 Gödel showed that the Continuum Hypothesis cannot be refuted in (first-order) ZFC. His argument was by an inner model consisting of the constructible universe of sets. By 1947 Gödel was convinced that it could also not be proved in ZFC (1947; 1995, 183). In a series of three papers in the early 1950s John Shepherdson carefully explored the properties of inner models of set theory and showed, in the third paper, that no inner model construction can establish that the negation of the continuum hypothesis is consistent with ZFC. In the second paper Shepherdson reworked Zermelo's categoricity theorem for ZFC as a theorem about inner models of NBG. I will discuss the significance of this reworking later.

Paul Cohen discovered the method of forcing as a means of constructing "outer" models of set theory, i.e. given a model of set theory, how to add sets to it, and in 1963 and 1964 published his proof that the continuum problem cannot be proved in ZFC (Cohen 1963, 1964). Eighty-five years after Cantor had formulated the Continuum Hypothesis, the lack of a solution seemed to be not a matter of its difficulty but of the now mathematically established fact that no known mathematics could solve it.

This was the first ever such result. In the preceding hundred years there had been a number of important independence results in mathematics, starting with the proof by construction of models that Euclid's fifth postulate is independent of the other axioms of Euclidean geometry, from the work of Riemann, Beltrami, Klein and others. Then had come the proof theoretic independence proof of the Gödel sentence. Now, however, for the first time, a problem whose solution was sought by mathematicians was shown to be unsolvable by all known means. There was consternation and confusion in the immediate aftermath of this result and some, including Paul Cohen (1967), considered that the continuum problem should now be considered not to be a genuine problem. This of course was not Gödel's view, who had anticipated the independence of CH and firmly held to its being a genuine problem, which he continued to attempt to solve (see Gödel 1970).

In 1965, in the aftermath of these developments, a conference on philosophy of mathematics took place in London in which the status of the continuum problem was much discussed. Those giving papers, chairing sessions, or simply attending included Kreisel, Bernays, Mostowski, Kalmár, Tarski, Carnap, Quine, Dummett, Kleene. Two of

the papers, by Andrzej Mostowski and Paul Bernays, included the phrase ‘recent results in set theory’ in their titles. The conference volume, edited by Imre Lakatos, captures the ferment of the occasion by including discussion as well as the papers presented.

Two of these papers, by Paul Bernays (1967) and Georg Kreisel (1967), advocated the idea that the categoricity of second-order ZFC establishes that the continuum problem remains a problem with a determinate answer despite the result that it cannot be decided by the first-order axioms of ZFC. This view met with great hostility from Mostowski, both in his own paper and in his contributions to the discussion, and in discussion by Kalmár and Bar-Hillel.

In his paper Kreisel adumbrates a crucial insight into the nature of mathematics and foundations of mathematics by focusing on the notion of “informal rigor.”⁷ It seems to me that philosophy of mathematics should pay much more attention to this notion than has been the case. One difficulty about this paper of Kreisel as a source or resource for such attention is that Kreisel does not discuss the notion so much as take it for granted. Kreisel begins his paper by observing the limits of formal rigor:

Formal rigor does *not* apply to the discovery or choice of formal rules nor of notions; neither of basic notions such as *set* in so-called classical mathematics, nor of technical notions such as *group* or *tensor product*. (Kreisel 1967, 138)

We can extract from Kreisel’s paper the following positive characterization of informal rigor:

⁷ Georg Kreisel, born in 1923, did an undergraduate degree in mathematics at Trinity College, Cambridge from 1940 to 1943. He became a mathematical logician by reading Volume II of Hilbert and Bernays (there were no mathematical logicians in Cambridge at that time with whom he could have studied). While an undergraduate he had regular discussions on philosophy of mathematics with Wittgenstein. According to Ray Monk (1990, 498), “in 1944 Wittgenstein shocked Rush Rhees by declaring Kreisel to be the most able philosopher he had ever met who was also a mathematician. ‘More able than Ramsey?’ Rhees asked. ‘Ramsey?!’ replied Wittgenstein. ‘Ramsey was a mathematician!’” At Kurt Gödel’s invitation Kreisel spent two years, 1955–1957, at the Institute for Advanced Studies in Princeton and a further year 1963–1964. While Kreisel’s work is always mathematically informed, the problems he has worked on are philosophically motivated and contain important philosophical insights, though these insights are generally implicit or in passing rather than the focus of any sustained discussion.

The “old fashioned” idea is that one obtains rules and definitions by analyzing intuitive notions and putting down their properties. This is certainly what mathematicians thought they were doing when defining length or area or, for that matter, logicians when finding rules of inference or axioms (properties) of mathematical structures such as the continuum ... What the “old fashioned” idea assumes is quite simply that the intuitive notions are *significant*, be it in the external world or in thought (and a *precise* formulation of what is significant in a subject is the result, not a starting point of research into that subject). Informal rigor wants (i) to make this analysis as precise as possible (with the means available), in particular to eliminate doubtful properties of the intuitive notions when drawing conclusions about them; and (ii) to extend this analysis, in particular not to leave undecided questions which can be decided by full use of evident properties of these intuitive notions. (138–139)

Kreisel cites Zermelo’s axiomatization of set theory as an example:

Zermelo’s analysis furnishes an instance of a rigorous *discovery of axioms* (for the notion of set) ... What one means here is that the intuitive notion of the cumulative type structure provides a coherent *source* of axioms; our understanding is sufficient to avoid an endless string of ambiguities to be resolved by further basic distinctions, like the distinction between abstract properties and sets of something ... Denying the (alleged) *bifurcation* or *multifurcation of our notion of set of the cumulative hierarchy* is nothing else but asserting the properties of our intuitive conception of the cumulative type structure. This does not deny the established fact that, in addition to this basic structure, there are also technically interesting non-standard models; cf. App. B, defined in terms of the basic structure. (144–145)

Note Kreisel’s use of the word ‘structure’ in this passage. In Appendix B to this paper, on “Standard and nonstandard models,” to which Kreisel refers in the last sentence of the above quotation, the notion of structure receives the following elucidation:

[i]f one thinks of the axioms as *conditions* on mathematical objects, i.e. on the structures which satisfy the axioms considered, these axioms make a selection *among* the basic objects; they do not tell us what the basic objects are. (165)

Kreisel gives four applications of his notion of informal rigor,

mostly following the “old fashioned” idea of pushing a bit farther than before the analysis of the intuitive notions considered. Section 1 concerns the difference between familiar independence results, e.g. of the axiom of parallels from the other axioms of geometry, on the one hand and the independence of the continuum hypothesis on the other; the difference is formulated in terms of higher order consequence. (139)

Kreisel cites the following examples of the usefulness for informal rigor of thinking in terms of second-order consequence:

The familiar classical structures (natural numbers with the successor relation, the continuum with a denumerable dense base etc.) are definable by *second* order axioms, as shown by Dedekind. Zermelo showed that his cumulative hierarchy up to ω or $\omega + \omega$ or $\omega + n$ (for fixed n) and other important ordinals is equally definable by second order formulae. (148)

Let $\mathcal{Z}$ be Zermelo’s axiom with the axiom of infinity and let CH be the (canonical) formulation of the continuum hypothesis ... As Zermelo pointed out (see above [referring to the passage just quoted]), if we use the current set-theoretic definition $Z(x)$ of the cumulative hierarchy, in any model of $\mathcal{Z}$, this formula Z defines a [level of the cumulative hierarchy] C_σ for a limit ordinal $\sigma > \omega$. Consequently we have $(\mathcal{Z} \vdash_2 \text{CH}) \vee (\mathcal{Z} \vdash_2 \sim \text{CH})$. (150)

Kreisel elucidates the nature of this second order determination (he does not use this word) of CH by contrast with the Axiom of Replacement and the Euclid’s fifth postulate.

In contrast to the example on CH above, Fraenkel's replacement axiom is not decided by Zermelo's axioms (because $\mathcal{Z}$ is satisfied by $C_{\omega+\omega}$ and Fraenkel's axiom is not); in particular it is independent of Zermelo's second order axioms while, by Cohen's proof, CH is only independent of the *first order schema* (associated with the axioms) of Zermelo-Fraenkel ... Secondly, it shows a *difference* between the independence of the axiom of parallels in geometry on the one hand and of CH in first order set theory [on the other]. In geometry (as formulated by Pasch or Hilbert) we also have a *second order* axiom, namely the axiom of continuity or Dedekind's section: *the parallel axiom is not even a second order consequence of this axiom*, i.e. it corresponds to Frenkel's axiom, not to CH. (150–151)

In order actually to solve the continuum problem a formalizable derivation from axioms, of the kind which Cohen and Gödel's results show not to exist from the first-order axioms of ZF, must be found. This means that new axioms are required.

[n]ew primitive notions, e.g. properties of natural numbers, which are *not* definable in the language of set theory ... may have to be taken seriously to decide CH; for, what is left out when one replaces the second order axiom by the schema, are precisely the properties which are not so definable. (152)

Bernays' paper in the conference volume is much shorter than Kreisel's (three and a half pages compared to 34 pages), so does not develop its ideas to the degree found in Kreisel's paper. Nonetheless, Bernays says enough to show that on the points cited above he is completely in agreement with Kreisel.

Bernays begins by noting that

[t]he results of Paul J. Cohen on the independence of the continuum hypothesis do not directly concern set theory itself, but rather the axiomatization of set theory; and not even Zermelo's original axiomatization, but a sharper axiomatization which allows of strict formalization. (Bernays 1967, 109)

Bernays is here referring to the difference between Zermelo's second-order formulation of Replacement and its use as an axiom schema in first-order ZF.

But whereas the Cohen procedure leads to non-standard models, we can, by the aforementioned device, state a kind of categoricity of the axioms of set theory, as was done, without formalization, by Zermelo in his *Grenzzahlen und Mengenbereiche*. Thus we see that the independence of the continuum hypothesis is essentially tied to the formalization of set theory. It is therefore a fact of a similar kind to the existence of non-standard models for formalized number theory. (110)

The "aforementioned device" of this quotation is the means by which "it is, for instance, possible to prove formally the categoricity of number theory" (1967, 109). Bernays spells out the connection to the continuum problem in the following terms, which formulate the idea that in Zermelo's second-order axiomatization the power of the continuum is determinate.

If the strictly formal methods in axiomatic set theory are transgressed by applying the schema of the selection-axiom and that of the replacement axiom with an *unrestricted* concept of predicate, then model theory shows that the power of the continuum must be the same for each model of the Zermelo-Fraenkel axioms. (111)

Bernays goes on immediately to stress that this fixity of the power of the continuum in no way gives us a solution to the continuum problem: "yet we are not able to determine by any of the known methods, what in fact is the power of the continuum" (111) and amplifies this point by observing that

[o]ur inability to deal successfully with the continuum problem is certainly connected with the circumstance that our explicit knowledge of the continuum is very restricted. We are not even able to define effectively a subset of the continuum which

can be shown to have the power of the second number class.
(111–112)

Though Bernays does not talk about informal rigor, he talks about the necessity for mathematics of *intuitive* proof, at which point he cites Kreisel:

According to an extreme form [of the “formalistic” doctrine], the significance of a mathematical theorem consists merely in the fact that it is found to be provable in an adopted formal deductive system. This view is in any case defective. For instance, even when we can derive a formula ‘for all x : $A(x)$ ’ in the adopted system, we do not thereby know that $A(x)$ really holds for every x , unless a consistency proof is given for the adopted system. But this then is an *intuitive* proof of a general number-theoretic theorem which must be understood *in the normal way* in order to yield the wanted result. This point has been repeatedly stressed by Georg Kreisel. (110)

In 1965, the same year in which the colloquium at which the preceding papers were presented took place, Paul Benacerraf published his paper “What Numbers Could Not Be.” Though it was explicitly in opposition to Frege (whose Julius Caesar problem is quoted as an epigraph) and espouses a form of structuralism, it does not go back to the structuralism of Hilbert that provoked Frege’s ire but rather the off-hand view of Poincaré cited above. Poincaré, who is not particularly famous for that viewpoint, is not mentioned but a connection with what the (generic) mathematician thinks is drawn prominently by the other epigraph of the paper, a long quotation from Richard M. Martin (1963, 3). In the passage quoted Martin claims, without citing any particular mathematicians, that “the attention of the mathematician focuses primarily upon mathematical structure, and his intellectual delight arises (in part) from seeing that a given theory exhibits such and such a structure ... the mathematician is satisfied so long as he has some ‘entities’ or ‘objects’ ... to work with, and he does not inquire into their inner character or ontological status.” He goes on to say “The philosophical logician [of whom Martin undoubtedly considered himself one], on the other hand, is more sensitive to matters of ontology and will be

especially interested in the kind or kinds of entities there are actually.” For Benacerraf this quotation is a peg on which to hang his declaration of allegiance to the mathematical rather than the philosophical viewpoint: “Martin goes on to point out (approvingly, I take it) that the philosopher is not satisfied with [the mathematician’s] limited view of things. He wants to know more and does ask the questions in which the mathematician professes no interest. I agree. He does. And mistakenly so” (Benacerraf 1965, 69).

Benacerraf begins by considering two different set-theoretic interpretations of the natural numbers (those of Zermelo and von Neumann) and arguing then that “if numbers are sets, they must be *particular* sets ... But if the number 3 is really one set rather than another, it must be possible to give some cogent reason for thinking so ... But there seems little to choose among the accounts ... any feature of an account that identifies 3 with a set is a superfluous one ... therefore 3, and its fellow numbers could not be sets at all” (1965, 62). Benacerraf then generalizes this argument, to conclude that “numbers could not be objects at all; for there is no more reason to identify any individual number with any one particular object than with any other (not already known to be a number)” (69).

This is what numbers could not be, namely particular objects. But this negative thesis also led Benacerraf to a positive one. “Any object can be the third element in some progression. What is peculiar to 3 is that it defines that role—not by being a paradigm of any objects, which plays it, but by representing the relation that any third member of progression bears to the rest of the progression. Arithmetic is therefore the science that elaborates the abstract structure that all progressions have in common merely in virtue of being progressions” (70).

This positive thesis explains the negative one. Benacerraf immediately follows the statement just quoted with the conclusion that arithmetic “is not a science concerned with particular objects—the numbers. The search for which independently identifiable particular objects the numbers really are (sets? Julius Caesars?) is a misguided one” (70).

The positive thesis depends centrally on the notion of “the abstract structure that all progressions have in common merely in virtue of being progressions.” In the three pages of “What numbers could not be” after introducing this notion, Benacerraf provides no useful

discussion of it (part of what he said there he later explicitly retracted, see Benacerraf 1996).

With Benacerraf's 1965 paper the notion of mathematics as the science of abstract structures entered the mainstream of philosophy of mathematics for philosophers. Ten years later Michael Resnik published a paper citing Benacerraf (1965) for the argument against Platonism in terms of invariance of mathematics under isomorphism and propounding structuralism as a philosophy of mathematics. "I now want to propose an account which reflects the central aspects of mathematical activity while avoiding the pitfalls of the previous views. According to this view mathematics studies *patterns* or *structures*. For want of a better term I will call the view *structuralism*" (to which Resnik appends a footnote: "Views similar to mine have been discussed by C. Parsons, P. Benacerraf, Piaget and O. Chateaubriand") (Resnik 1975, 33). This is the first place, so far as I am aware, that "structuralism" was used as a label for a philosophy of mathematics. Resnik understands the notion of (mathematical) structure in terms of pattern recognition: "I see mathematicians as studying patterns or structures *qua* abstract entities. I picture the standard mathematical objects—numbers, functions, even sets and vectors—as positions in patterns or structures. (I will use the terms 'pattern' and 'structure' more or less interchangeably)" (Resnik 1988, 405). I find Resnik's formulation of the notion of structure empty.

The insight that mathematics is the study of structures is not by itself a philosophy of mathematics, and there are now a number of views each more or less incompatible with the others whose proponents call their account a structuralist philosophy of mathematics. At the heart of any would be structuralist philosophy of mathematics must be an answer to the question: What is a mathematical structure? The answer to that question and the resulting formulation of structuralism adumbrated in this paper differs at key points from the various current views that adopt this label. I shall spell out some of these differences in section 4.2.

In the 1980s Stewart Shapiro began to propound a form of structuralism. Shapiro acknowledges Resnik's influence, but his form of structuralism is quite different in terms of understanding what a mathematical structure is, which instead links up with the development stemming from Dedekind, Hilbert, and Zermelo and continuing in the papers of Kreisel and Bernays cited above. This approach is much

more illuminating, but the question what a mathematical structure is was not solved. The difficulty is that structures are made up of objects. But the whole point of structuralism is to account for mathematical objects in terms of structures. In 1989 Geoffrey Hellman published a book, *Mathematics without Numbers*, in which he attempted to finesse this problem by treating the objects that structures are made up out of as existing only possibly rather than actually. This is a distinction without a difference as far as the objects of mathematics are concerned.

In the next section I offer an account of the notion of mathematical structure that I consider constitutes the basis for structuralism as a philosophy of mathematics.

4 Particular mathematical structures

4.1 *Particular vs. general*

Mathematicians study two sorts of structures, which I shall call *particular structures* and *general structures*. The distinction is marked by use of the definite and indefinite articles. We speak of *the* natural numbers and *a* group. Particular structures include the natural numbers, the Euclidean plane, the real numbers. General structures include groups, rings, fields, metric spaces, topologies. The particularity of a particular structure consists in the fact that all its exemplars are isomorphic to each other. The generality of a general structure consists in the fact that its various exemplars need not be, and in general are not, isomorphic to each other.

The exemplars of a given general structure may all be exemplars of another general structure, e.g. every ring is a group, but the exemplars of any general structure are particular structures. For example, the permutations on three letters under composition is a group, also the integers under addition (but not under multiplication) and the rational numbers under addition and under multiplication, also the distance-preserving transformations of the Euclidean plane, and for each natural number $n > 1$, the natural numbers under addition mod n . A particular structure may exemplify several general structures, e.g. the real numbers are a metric space with respect to the function $|x - y|$ and a topological space whose basic open sets are the open intervals determined by the $<$ -relation, and the reals may be endowed with infinitely many different topologies.

Initially, mathematics was all about using and studying particular structures (the natural numbers, later the integers, the Euclidean plane, Euclidean three-space, the rationals, later the reals, still later the complex numbers). The recognition and study of general structures began in the first half of the 19th century, with the work of Galois and others on the solution of polynomial equations, which led to the notion of a group. By the second half of the 20th century general structures were paramount in mathematics, a viewpoint proclaimed by Nicholas Bourbaki in his manifesto “L’architecture des mathématiques.”⁸ Bourbaki talks about “structures mères,” translated literally as “mother-structures” (Bourbaki 1948, English trans. 228), which are algebraic structure, order structure, and topological structure.

One virtue of general structures is their generality. A theorem about groups establishes something true about addition on the integers, composition of transformations of the Euclidean plane, permutation of the roots of a polynomial over the reals. A proof that in every group the inverse of an element is unique establishes that negation is a well-defined function on the integers and at the same time that for each matrix its inverse matrix is unique. Also, a theorem about groups establishes something about other general structures, e.g. something that is true in every ring is true in every field. A proof in the axiomatic theory of a general structure establishes something true about *every* particular structure satisfying those axioms, e.g. each particular group. But also, and this is a major reason for Bourbaki’s enthusiasm for general structures, such a proof will also prove something about any other general structure which includes among its axioms the axioms of that structure, e.g. topological groups immediately inherit all the theorems of group theory and of topology. Also, a general structure gives rise to

⁸ ‘Nicholas Bourbaki’ was a pseudonym adopted in the 1930s by a group of young French mathematicians united in their conception of how modern mathematics should be understood and developed. Under this collective name they published treatises expounding each branch of mathematics in the manner of their conception of mathematics and edited their *Seminaire Bourbaki*. Their paper “L’architecture des mathématiques” [The architecture of mathematics] (Bourbaki 1948), was a kind of manifesto of their viewpoint, written by Jean Dieudonné, a leading member and sometimes spokesperson of the group (see Corry 1998, 158).

a particular structure G , namely the category of G s, e.g. the category of groups.

More important than the generality of general structures is their cross-fertilization with other structures, by which, e.g. complex analysis becomes a tool for number theory. Out of the “structures mères” are formulated

structures which might be called multiple structures. They involve two or more of the great mother-structures simultaneously not in simple juxtaposition (which would not produce anything new), but combined organically by one or more axioms which set up a connection between them. Thus, one has topological algebra. This is a study of structures in which occur, at the same time, one or more laws of composition and a topology, connected by the condition that the algebraic operations be (for the topology under consideration) continuous functions of the elements on which they operate. Not less important is algebraic topology, in which certain sets of points in space, defined by topological properties (simplexes, cycles, etc.) are themselves taken as elements on which laws of composition operate. (Bourbaki 1948, English trans. 229)

This cross-fertilization between general structures is crucial to the development of contemporary mathematics.

Farther along we come finally to the theories properly called particular. In these the elements of the sets under consideration, which, in the general structure have remained entirely indeterminate, obtain a more definitely characterized individuality. At this point we merge with the theories of classical mathematics, the analysis of functions of real or complex variable, differential geometry, algebraic geometry, theory of numbers. But they have no longer their former autonomy. They have become crossroads, where several more general mathematical structures meet and react upon one another. (1948, 229)

A supreme example of this cross-fertilization in modern mathematics is the proof of Fermat’s Last Theorem, which concerns just the structure

of the natural numbers, from the Shimura-Taniyama-Weil conjecture concerning elliptic curves.⁹

Bourbaki is contemptuous of axiomatic theories of the particular structures and considers that the triviality of axioms which pick out only one structure explains why too many mathematicians at that time dismissed the axiomatic method.

Many of the [mathematicians seriously opposed to the development of mathematics in axiomatic theories] have been unwilling for a long time to see in axiomatics anything else than futile logical hairsplitting not capable of fructifying any theory whatever. This critical attitude can probably be accounted for by a purely historical accident. The first axiomatic treatments and those which caused the greatest stir (those of arithmetic by Dedekind and Peano, those of Euclidean geometry by Hilbert) dealt with univalent theories, i.e. theories which are entirely determined by their complete system of axioms; for this reason they could not be applied to any theory except the one form which they have been extracted (quite contrary to what we have seen, for instance, for the theory of groups). If the same had been true for all other structures, the reproach of sterility brought against the axiomatic method would have been fully justified. (1948, 230)

Bourbaki extols the axiomatic method as the means by which to study general structures, contrasting it with the axiomatization of particular structures. When introducing the group axioms in “*L’architecture des mathématiques*” he declares that “there is no longer any connection

⁹ “The Shimura-Taniyama-Weil conjecture relates elliptic curves (cubic equations in two variables of the form $y^2 = x^3 + ax + b$, where a and b are rational numbers) and modular forms, objects arising as part of an ostensibly different circle of ideas. An elliptic curve E can be made into an abelian group in a natural way after adjoining to it an extra ‘solution at infinity’ that plays the role of the identity element. This is what makes elliptic curves worthy of special study, for they alone, among all projective curves (equations in two variables, compactified by the adjunction of suitable points at infinity) are endowed with such a natural group law. If one views solutions geometrically as points in the $(x; y)$ -plane, the group operation consists in connecting two points on the curve by a straight line, finding the third point of intersection of the line with the curve and reflecting the resulting point about the x -axis” (Darmon 1999, 1397).

between this interpretation of the word ‘axiom’ and its traditional meaning of *evident truth*” (225, fn. **). For Bourbaki this modern viewpoint supersedes the old, benighted, viewpoint.

However much this may be true of the 20th century development of mathematics, it is not true from the point of view of philosophy of mathematics. The reality of mathematics, including its modern development, rests ultimately upon the reality of the particular structures of mathematics. A general structure for which no particular structures are exemplars is vacuous. From a philosophical point of view the nature and reality of mathematics comes down to the existence and nature of particular structures, and for them, *pace* Bourbaki, their axioms are “evident truths.”

In the case, for example, of Dedekind’s axiomatizations of the structure of the natural numbers (1888), and of the continuum (the real numbers) (1872), he was looking for, and found, “evident truths” by which to characterize these structures. The same is true of Euclidean geometry, though some confusion surrounds this case. The view is sometimes, or even often, taken that the discovery of non-Euclidean geometry showed that the axioms of Euclidean geometry are not to be thought of as “evident truths” because they are not, as it turns out, truths at all. This viewpoint is arrived at from either of two mistaken considerations. One is based on the fact that relativistic physics shows that the geometry of the (physical) universe is non-Euclidean, so Euclid’s fifth postulate is not, as it turns out, true. The other takes it that the fifth postulate holds in Euclidean geometry and not in non-Euclidean geometries. Since no sentence and its negation can both be true, the fifth postulate is, on this other viewpoint, neither true nor false. There is no true geometry. It is, of course, correct that there is no *one* true geometry. But each geometry, Euclidean and each non-Euclidean geometry, is a particular geometry, the axioms of which are evident truths for that particular structure, just as the Dedekind axioms are evident truths for the structure of the natural numbers.

Bourbaki offers an account of general structures which Leo Corry (1992, esp. 329, 340–342) and others have argued is not as mathematically fruitful and powerful as category theory. Be that as it may, and crucial as the study of general structures is for contemporary mathematics, it is the nature of particular structures that is crucial to our project. By failing to distinguish between these two kinds of structure

in mathematics, those who argue for the foundational primacy of category theory as theory of structure are at cross purposes with structuralism as a philosophy of mathematics in the sense that is under discussion in this paper.

4.2 Three untenable accounts of what particular mathematical structures are

Before offering a positive account of what particular mathematical structures are, I want to clear away three other accounts: (1) the notion of structure from model theory (the basis of a rich mathematical theory but no good for these purposes), (2) Shapiro's axiomatic theory of structures, (3) modal structuralism.

(1) The model-theoretic notion of structure: A mathematically natural answer to the question, "What is a particular mathematical structure?" is that a particular mathematical structure consists of a domain of objects and relations between those objects, where an n -ary relation over a domain D is a subset of the n -fold Cartesian product of D with itself. This is the notion of structure that is subject of mathematical investigation in model theory. If this is what a mathematical structure is then we cannot say that it is structures rather than objects that constitute the locus of mathematical reality since structures are constructed from objects by set-theoretic constructions. Rejection of *in re* structuralism means that the highly articulated mathematical theory of structures stemming from model theory, in the work of Tarski and others, cannot be of use for these philosophical purposes. The point is that the model-theoretic notion of structure takes as its starting point a domain of objects and is a construction (definition) within set theory with urelemente, or within pure set theory. Insofar as the notion of mathematical object is philosophically problematic, appeal to this account begs the question.

(2) Stewart Shapiro's theory of structures (1997, chapter 3, section 4): Shapiro's aim is to "axiomatize the notion of structure directly. The envisioned theory has variables that range over structures and thus a quantifier 'all structures'"(93). Certainly we have to answer the question, What is a mathematical structure? Whether there can be a mathematical theory of structure is not clear. Shapiro notes that "because isomorphism and structure-equivalence are equivalence relations, one

can informally take a structure to be an isomorphism type or a structure-equivalence type. So construed, a structure is an equivalence class in the set-theoretic hierarchy. Notice, however, that each nonempty ‘structure’ is a proper class, and so it is not in the set-theoretic hierarchy” (92). As various commentators have pointed out, the Burali-Forti paradox threatens attempts to find more manageable objects as structures.¹⁰ However these problems are dealt with, the result will be a mathematical theory of mathematical structures, which will then be a piece of mathematics that itself must be accounted for (similar to the problem with the model-theoretic account of structure). Either it’s viciously circular or the start of an infinite regress.¹¹

In any case Shapiro at this point is not opting for a mathematical theory of structure that provides an explicit account of what things mathematical structures are. Rather, as noted above, his attempt is to “axiomatize the notion of structure directly.” What he offers as axioms of structures do no such thing. They have the character of Euclid’s definitions rather than his axioms, e.g. “A point is that which has no part. A line is breadthless length. The ends of a line are points. A straight line is a line which lies evenly with the points on itself.” And so on. This might seem a distinguished pedigree, but for all the genius of Euclid’s axiomatization of geometry, this aspect has been recognized as misguided. The trouble for Shapiro’s theory of structures is that there is no possibility of anyone producing a re-axiomatization as a *Grundlagen der Struktur* corresponding to what Hilbert did for Euclidean geometry. Thus for example Shapiro’s Axiom “Infinity” (93) helps itself to the notion of infinity: “There is at least one structure that has an infinite

¹⁰ Cf. Linnebo (2008, 75): “I can now define *isomorphism types* of relations by the following abstraction principle: $\overline{R} = \overline{R'} \leftrightarrow R \cong R'$. But one has to be careful here, for without any restriction, $\overline{R} = \overline{R'} \leftrightarrow R \cong R'$ leads straight to paradox,” to which is appended the footnote, “It will allow us to derive Burali-Forti’s paradox: see Hazen 1985, 253–254.”

¹¹ Crispin Wright offered the suggestion in discussion (when I presented some of this material in a seminar at St Andrews) that there is an alternative to these two possibilities, namely an infinite hierarchy. This observation is certainly true, and shows that there is no knock-down argument against the possibility of a mathematical theory of structures in the sense required. On the other hand, Shapiro’s theory of structures, the target of this discussion, is not ramified in this way. Whether there may be a mathematically cogent theory of structures that would constitute a basis for a structuralist understanding of the nature of mathematics is left open.

number of places.” Of course axiom systems do that, e.g. it’s an axiom of Euclidean geometry that “two points determine a line,” and the axiom system as a whole determines what points and lines are. But Shapiro’s Axiom of Infinity, along with his other axioms, does *not determine* any notion. It is *using* the notion as understood. Similarly, and even more problematically, his axiom “Coherence” (95) helps itself to the notion of a “coherent” formula (in a second-order language): “If Φ is a coherent formula in a second-order language, then there is a structure that satisfies Φ .” Introducing this axiom, Shapiro notes that “the main principle behind structuralism is that any coherent theory characterizes a structure, or a class of structures” (95). Quite apart from the pressing question, about which I will say something below, whether it makes any sense to declare this principle an “axiom,” there is the fundamental question, what makes it true?

If a theory is expressed in a first-order language and coherence is construed as formal consistency, this principle is a theorem of logic, i.e. the Gödel completeness theorem, which tells us that any first-order theory that is formally consistent has a model. But the models proved to exist are not, in general, the mathematical structures that structuralism is about. For one thing they are all countable. And even in cases where the “intended structure” is countable, e.g. the natural numbers, the resulting structure may be non-standard (e.g. a structure in which there are infinitely large elements). As Shapiro and others have long noted, the language in which to articulate our understanding of particular mathematical structures is second-order, and Shapiro notes that in this case we cannot understand the notion of coherence of a mathematical theory in terms of formal consistency: “Notice, for now, that because we are using a second-order language, simple (proof-theoretic) consistency is not sufficient to guarantee that a theory describes a structure or class of structures” (95). (For example, if we add to Dedekind’s second-order axioms for arithmetic the infinite set of sentences $a \neq \bar{n}$ for each numeral $\bar{n}$, the resulting theory is incoherent, i.e. there is no structure that it is about, but it is not formally inconsistent, since any derivation from that theory uses only finitely many of its axioms, so all formulas in the derivation, including the conclusion, can be interpreted in the structure of the natural numbers, which tells us that the conclusion cannot be a contradiction.) So we are back to the question how it is that the coherence of a mathematical theory means that it characterizes a structure.

Seemingly, Shapiro considers that this connection between coherence of a theory and the structure that the theory characterizes can itself be characterized axiomatically. Insofar as this is what he is saying, it seems to me misguided. Understanding this relationship belongs to philosophy of mathematics and *not* to mathematics, and so *cannot* be characterized axiomatically. Shapiro remarks, “We can, of course, add an axiom that, say, second-order ZFC is coherent, and thus conclude that there is a structure the size of an inaccessible cardinal” (95). The idea that such an axiom can be given a mathematically meaningful formulation—in what language could it be expressed?—let alone that we can achieve anything by such an axiom, seems to me whistling in the dark. If we want to conclude that there is a structure the size of an inaccessible cardinal, we do so by *showing* that second-order ZFC is coherent, not by *postulating* that it is. How do we show that ZFC is coherent? By developing our mathematical understanding of the subject matter of this theory through informal rigor, as was done by Zermelo (1930) and Shepherdson (1951–1953). We achieve understanding of the notion of mathematical structure not by axiomatizing the notion but by reflecting on the development of mathematical practice by which particular mathematical structures come to be understood, the natural numbers, the Euclidean plane, the real numbers, etc.

How do we know that such structures exist? The question is likely to be construed in such a way that it is a bad question. There is nothing we can do to establish that particular mathematical structures exist apart from articulating a coherent conception of such a particular structure. If this is what Shapiro means when he says, “If Φ is a coherent formula in a second-order language, then there is a structure that satisfies Φ ,” I must apologize for my previous misunderstanding. However, the choice of language suggests this is not what he means, in particular that we are meant to understand the notion of something being a structure that satisfies Φ as different from the notion of Φ being a coherent theory.

For Shapiro the case of set theory is singular:

structure theory, as I conceive it, is about as rich as set theory. It has to be if set theory itself is to be accommodated as a branch of mathematics. In a sense, set theory and the envisioned structure theory are notational variants of each other. (1997, 96)

In my view set theory, because of its universality, is an *experimentum crucis* for establishing a structuralist understanding of mathematics but, crucially, it is one case among others and *not* singular. Shapiro's view that set theory and what he envisages as structure theory "are notational variants of each other" constitutes a collapse into *in re* structuralism.

(3) Hellman's modal structuralism (1989): The point can be put briefly. Prefixing the symbol $\Diamond$ to the second-order axioms of a particular mathematical structure is an empty gesture. There is no difference between possible and actual existence for objects of mathematics. For further discussion of Hellman's modal structuralism, see pages 40–41.

4.3 *How to say what particular mathematical structures are; the reality of mathematics and the unreality of mathematical objects*

However natural and mathematically fruitful the model-theoretic/set-theoretic notion of structure is, a different understanding is required for our philosophical purposes. The beginning of such an understanding is the point that structures are characterized by axioms. This reflects and also explains the fact that the axiomatic method is intrinsic to mathematics.¹²

I need to make use of a distinction between *abstract* and *concrete* structures within the topic of *particular structures*. It is roughly the difference between *type* and *token*, a distinction that Stewart Shapiro articulates as follows.

Because the same structure can be exemplified by more than one system, a structure is a one-over-many. Entities like this have received their share of philosophical attention throughout the ages. The traditional exemplar of one-over-many is a universal, a property, or a Form. In more recent philosophy, there is the type/

¹² That said, in the development of each branch of mathematics there is, in general, a period of pre-axiomatic development. Geometry was a branch of Greek mathematics for several hundred years before its axiomatization by Euclid. The axiomatization of geometry was an extraordinary piece of mathematics. It would be quite wrong to think (as sometimes is said) that Euclid was *merely* systematizing an existing practice of mathematics. The pre-axiomatic development of number theory lasted for several thousand years. The pre-axiomatic study of set theory lasted for forty years, throughout the work of Georg Cantor, from around 1865 until Zermelo's axiomatization of set theory in 1908.

token dichotomy. In philosophical jargon, one says that several tokens *have* a particular type, or *share* a particular type; and we say that an object *has* a universal or, as Plato put, an object *has a share of*, or *participates in* a Form. As defined above, a structure is a pattern, the form of a system. A system, in turn, is a collection of related objects. Thus, structure is to structured as pattern is to patterned, as universal is to subsumed particular, as type is to token. (Shapiro 1997, 84)

It is not contact with the natural numbers that gave rise to our grasp of the structure of the natural numbers. Our starting point is grasp of that structure through our use of suitable concepts.

Note that this is the point of the remark by Poincare quoted earlier. He is talking about particular rather than general structures, and making the point that particular structures are not constituted by particular objects.

Even if we take pure set theory as given and can thus (relative to our grasp of set theory) find a bunch of objects to exemplify the structure of the natural numbers, the only way we can do this is by deploying our grasp of the (abstract) structure of the natural numbers.

Another distinction I need to draw is between the model-theoretic notion of structure as defined by set theory with urelemente (non-sets) and structures defined in pure set theory. It is the latter which constitutes set theory as a foundation of mathematics.

Consider those forms of *in re*, or reductive, structuralism, that treat pure sets as “ontological atoms.” Even if we take pure set theory as given, so that we can take the structure of the natural numbers as exemplified in set theory (note that “exemplify” is Shapiro’s word in the passage I quoted above), the only way we can do this is by deploying our grasp of the (abstract) structure of the natural numbers to find this exemplification, by suitably describing, such a collection of sets and a successor function.

One of the ways in which current forms of so called *ante rem* structuralism goes wrong is that it goes from *ante rem* structures to objects of mathematics, as places in the structures. The impression that in this way an ontology of mathematical objects has been explained is an illusion. One way to see this seeming explanation to be an illusion is that the explanatory order from *ante rem* structure to mathematical objects

as places in structures does not (of course) reflect any temporal order. So when all is said and done in the explaining that we think we are doing, we have revealed not only the existence of *ante rem* structures but also the existence of all the mathematical objects that anyone ever thought there were. This should cause alarm bells to ring. The incoherence of believing that there are all the mathematical objects that each mathematical theory appears to be about is in no way resolved by this story. The coherence of talking about structures is all that we have got in explaining what we are doing when we do mathematics. The coherence of our understanding of the structure of the real continuum does not translate into an account of each individual real number. Of course not. Very crudely, cardinality considerations alone tell us that there is no such account of individual real numbers. We always understand our mathematics by understanding the relevant structure.

Consistency and categoricity are the two criteria that determine whether we have succeeded in identifying a mathematical structure. Without consistency categoricity is vacuous. Consistency without categoricity means that we have not picked out a particular structure. We might not want to have picked out a particular structure (cf. Bourbaki), but if we do want to, categoricity is what is required.

The requirement of consistency is sufficient for structure, with categoricity determining whether that structure is general or particular.

To understand the notion of a particular mathematical structure we must work from examples of particular structures, mathematically important and in many cases studied over centuries, e.g. the structure of the natural numbers given, for example, by its second-order characterization in the language of 0 and the structure of the reals, but new particular structures are added to mathematics as it progresses, e.g. the p -adic numbers for each prime p . The key elements of the account are *categoricity* and *informal rigor* in the sense of Kreisel (1967). Categoricity gives us the particularity of the particular structure. Informal rigor is how we know what we are talking about. I also want to argue that the categoricity results for e.g. theory of the natural numbers and theory of sets show that these axiomatizations capture *fully* the intended notion. Such an account faces various challenges.

Some have claimed, e.g. Juuko Keränen (2001) and Fraser MacBride (2005) that the existence of particular structures that admit non-trivial automorphism, such as the complex numbers under

conjugation, refute structuralism. Such claims are mistaken. What force they appear to have is down to the failure, e.g. in Stewart Shapiro's formulation of what he calls *ante rem* structuralism, to understand that there are ultimately only structures and no objects. If one takes the existence of *ante rem* structures to be a basis for explaining the existence of the objects of mathematics as positions in the particular structure, structures with a non-trivial automorphism create a problem. But the idea that one is explaining the existence of mathematical objects is already wrong, and these arguments to show that in the case of particular structures with automorphism we do not have an account of the existence of the objects of that theory may be a problem for forms of structuralism that claim to be saying what the objects of mathematics are, but not for the structuralist account of mathematics that I am here advancing.

The greatest challenge structuralist philosophies of mathematics face is to answer the question, What *is* a mathematical structure? Among philosophers of mathematics who count themselves structuralist there is no unanimity on how to answer this question. Stewart Shapiro, eminent among structuralists, has stressed the contrast between considering a particular structure as existing in virtue of objects out of which it is composed (e.g. individual natural numbers) or as existing independently of the existence of objects that are configured in it. To mark this contrast he has adopted terminology that has come into philosophers' English from the Latin of medieval philosophy, *in re* versus *ante rem*, as in the *Oxford English Dictionary*:

in re: in reality; (of universals) dependent for their existence on the existence of the particulars that instantiate them, as Aristotle held; having real or objective, not merely mental, existence, but not separately from particulars

ante rem: prior to the existence of a particular or physical thing; specifically of the philosophical theory which holds that the universal is logically prior to, i.e. capable of existing independently of, the particular.

Shapiro carries over this terminology concerning the existence of universals to the existence of mathematical structures.

Insofar as we wish to avoid the (in my view) hopeless idea that we can make sense of mathematical objects existing individually, i.e. independently of a structure in which they occur, the first of these possibilities is ruled out. General structures are given by their axioms, which are stipulative (a group is any set with binary operation satisfying the axioms of group theory). Particular structures are also given by their axioms, but not stipulatively, rather by the mathematical fact that any two models of those axioms are isomorphic. For a set of axioms to characterize an infinite structure categorically full second-order quantification is required. I will discuss our understanding of second-order quantification later.

In seeming agreement with Shapiro I consider that structuralism must be understood as *ante rem* rather than *in re*, though for Shapiro the boundary between the two positions is not absolutely sharp. After all, structures are always made up of objects, even if the structure is (in some sense) prior to the objects. But that is just the problem. What sense could that be? The whole virtue of structuralism as a philosophy of mathematics, in my view, is to free ourselves from the hopeless idea that the individual particular objects of mathematics exist. They may do, for all we know, but if they do they must be like Kantian “things in themselves,” about which we can know nothing—except their structural properties.

Second-order quantification is needed to express understanding of mathematical structures. The test of when we have succeeded is categoricity. Deduction is *not* the point. Insofar as second-order logic is used deductively it is indistinguishable in expressive power from first-order logic (cf. Väänänen 2001), i.e. there is no formal difference between full second-order logic and second-order logic where the second-order quantifiers range just over the definable subsets of the domain of individuals. The crucial requirement in appealing to full second-order logic is *informal rigor*.

This view is well-tested by considering the case of set theory, in terms of the quasi-categoricity of ZF^2 . The case of set theory is particularly testing because set theory is often taken to be the framework within which all these results are to be formulated. So if there is going to be any vicious circularity in this way of explaining things it will show up here.

What is it for a set of second-order sentences S to be a categorical characterization of a particular structure? We say that for any two

structures A_1, A_2 in which S are true there is an isomorphism between A_1 and A_2 . But this criterion can not be used without explanation if we are trying to explain the notion of mathematical structure, since it requires the notion of *structure* to give this definition while the very notion we are attempting to explicate is that of mathematical structure.

Note, however, that the notion of categorical characterization of a structure does not itself require the existence of structures, since it says for any two structures that satisfies these axioms there exists an isomorphism between them. The existential claim for the isomorphism between the two structures is relative to the existence of the two structures; the isomorphism can be seen as made from the two structures (though it requires some set theory to do this). However, if this quantification over structures is vacuous, i.e. there are *no* such structures, then the categoricity theorem is vacuous, and every sentences in the language is a logical consequences of the theory.

There is an answer to this charge of circularity. The notion of structure that enters into the above definition is the ordinary (model theoretic) one (with a domain D and for a given similarity type subsets of D^m , etc.). No use of second-order language is required for this notion. Given this general, model-theoretic notion of structure, we can stipulate what is a *particular* structure, by which we are able to answer the philosophically (but also mathematically) motivated questions, “What is mathematics about?” and “In what does the reality of mathematics consist?”

Such a reply invites the following renewed challenge; How can you take it that the model-theoretic notion of structure is understood when ultimately that notion depends upon set theory, through its formulation in terms of an arbitrary domain and Cartesian products of the domain, etc? Yet you seek to account for the subject matter of set theory as a particular structure characterized in these terms that rely upon the model theoretic—which is to say ultimately set theoretic—notation of structure.

In some way this is my view. It goes with my rejection of *in re* structuralism, and my insistence that the Tarskian notion of structure cannot provide—is not a notion of structure that constitutes—a basis for structuralism as a philosophy of mathematics. But how then to avoid the danger of vacuity pointed out by Charles Parsons (1990, 310), Geoffrey Hellman (1989, 26), and others, that if there is no

structure in which a set of axioms hold then all sentences are its logical consequences, i.e. unless we know that *there exists* a structure in which Dedekind's second-order axioms for the natural numbers are true, then every sentence in that language is a logical consequence of those axioms?

I am attempting to avail myself of specific proofs of categoricity, e.g. of the theory of natural numbers, the real numbers, etc., i.e. that we have clear instances of the existence of particular structures. But the very claim that a specific theory is categorical involves the notion of structure. Let S be a specific (2nd order) theory (e.g. Dedekind arithmetic). The claim that S is categorical is the following: For every structure $\mathcal{A}$ and every structure $\mathcal{B}$, if $\mathcal{A} \models S$ and $\mathcal{B} \models S$ then there exists an isomorphism between $\mathcal{A}$ and $\mathcal{B}$. The notion of structure that is meant here is the set-theoretic one from 20th century logic, a domain of first-order quantification, D , and relations and functions as subsets of the n -fold Cartesian products of D for various natural numbers n . This looks like dependence on *in re* structuralism.¹³

There are two problems: (1) One is the universal quantification over structures, where the notion of structure is set theoretic, and so *in re*. (2) The other problem is that these universal quantifications must be non-vacuous. If not, the ostensible characterization of a particular structure is vacuous. The condition of non-vacuity is ontological, i.e. we must know that there exists such a structure. I call this the *exemplification problem*.

In respect of the first problem, the question of circularity appears to be particularly acute. We need set theory as the basis for quantifying over *all structures*. And yet set theory is one of the theories which picks out a particular (or in this specific case, quasi-particular) structure. The proper response, it seems to me, is to say that sets are not ontological atoms. Set theory is foundational in the sense that the language of set theory is pervasive in mathematics, but it is not true that everything in mathematics is made up of sets. (We must get completely away from Fregean foundationalism.) Set theory is to be understood structurally as much as all other mathematics.

¹³ The reference to natural numbers does not constitute a circularity since each claim is with respect to a specific theory so the similarity-type of the relevant structures is fixed and there is no quantification over the natural numbers in this formulation.

Concerning the exemplification problem, first note how not to deal with it, namely as Dedekind did, with his Theorem 66—though in spirit that is the right response. There is no way that the existence of any particular structure can be proved. But we can *know* that there is such a structure, as in the case of natural numbers. Fermat, for example, certainly knew that there is such a structure. Zermelo had the right idea when he said of his Axiom of Infinity that it was “essentially due to Dedekind,” and cites section 66 of Dedekind 1888, while at the same time noting that “the ‘proof’ that Dedekind there attempts to give of this principle cannot be satisfactory” (Zermelo 1908, 204).

In advocating an *ante rem* over *in re* notion of structure I am not opposed to abstract entities *per se*, indeed I accept and must insist upon the existence of various abstract entities (I am not a nominalist). I accept, for example, that works of fiction exist as abstract entities. So why not also numbers and sets as abstract entities? The point is that an individual natural number, e.g. 90745943887872000, exists only through its determination by the categorical theory of the structure of the natural numbers. It has no independent existence.¹⁴ Analogously, Shakespeare’s *The Tempest* exists as an abstract object (after he wrote it), but the character Prospero has no existence independent of that play. Even so, this is existence of a sort, namely derivative or dependent existence. But that is all it is.

I think it is strictly correct to say that mathematics is the study of structures only and there are no mathematical objects. It is a refinement of this point to say that the objects which apparently make up a particular mathematical structure can be taken to exist dependently on that structure. For a careful and perceptive discussion of structuralism and dependent existence of mathematical objects see Linnebo 2008. I fully agree with a great deal in that paper but disagree with its treatment of set theory (72–73), which goes wrong, it seems to me, in its unarticulated assumption that we can talk about sets without doing set theory.

¹⁴ This is not to say that individual natural numbers do not come into other mathematics than that of the structure of the natural numbers. Of course they do. (The number in the previous sentence, for example, is the order of one of the finite sporadic groups.) But it is only within the structure of the natural numbers, or structures into which the structure of the natural numbers is embedded, e.g. the system of base 10 numerals, that natural numbers as such occur.

The basis of mathematics is conceptual and epistemological, not ontological, and understanding particular mathematical structures is prior to axiomatic characterization. When such a resulting axiomatization is categorical, a particular mathematical structure is established. Particular mathematical structures are not mathematical objects. They are characterizations. Initially I took the view that at any given stage in the development of mathematics there are only finitely many particular mathematical structures, namely those that have been established by mathematicians up to that time. However, this is clearly wrong. While indeed there are up to any given moment of course only finitely many theorems establishing categorical characterizations of structures, e.g. of the natural numbers, the real and complex numbers, the Euclidean plane, the cumulative hierarchy of sets up to a particular ordinal, one such theorem may establish categorical characterization of infinitely many particular substructures. For example, having established the categoricity of the structure of the natural numbers, this result can be refined so that it can be seen also to establish the categoricity of each structure consisting of a particular natural number (this fact was pointed out to me by Marcus Rossberg). There are also cases where the categoricity of infinitely many particular structures that are not all substructures of a given particular structure can be established, e.g. for each prime number p , the structure of the p -adic real numbers. Crispin Wright has suggested to me that for each ordinal α , the structure of α is a particular structure. This is a more radical point than the case of the finite ordinals or the p -adic reals, where each of those structure can, in principle, be given an explicit categorical characterization. In the case of ordinals, explicit characterizations are bounded by Church-Kleene ω_1 , when ordinal notations run out. Nonetheless, it is provable that for each ordinal α , α is a particular mathematical structure.

It might seem that this last point threatens to collapse the notion of particular mathematical structure to the model-theoretic notion of structure. However, this is not the case. Consider, for example, the theorem that there are continuum many pairwise non-isomorphic countable models of first-order arithmetic. These countable non-standard models of first-order arithmetic are not particular structures in the sense at issue here. Their existence is inferred from the compactness theorem. There is no categorical characterization of countable non-standard models of arithmetic in their particularity.

How is model theory, as a branch of mathematics, to be understood on the understanding of the nature of mathematics being propounded here? Model theory proceeds, as does all mathematics, rigorously and informally, but its practice is not far from the theory within which it can be formalized, namely set theory. A structure in model theory is a construction in set theory with urelemente, the elements of the domain of the model being among the urelemente of the set theory. As remarked above, the development of set theory has shown that all structures of mathematics occur to within isomorphism in the cumulative hierarchy of pure sets, i.e. the theory of sets without urelemente. In this way the existence of the structures of model theory, as in the example cited above, are part of the development of pure set theory. To the extent that the truth of set theory can be accounted for in the structuralist terms set out in this paper, the theory of the structures of model theory is thereby also accounted for.

If the mathematical community at some stage in the development of mathematics has succeeded in becoming (informally) clear about a particular mathematical structure, this clarity can be made mathematically exact. Of course by the general theorems that establish first-order languages as incapable of characterizing infinite structures the mathematical specification of the structure about which we are clear will be in a higher-order language, usually by means of a full second-order language. Why must there *be* such a characterization? Answer: if the clarity is genuine, there must be a way to articulate it precisely. If there is no such way, the seeming clarity must be illusory. Such a claim is of the character as the Church-Turing thesis, for every apparently algorithmic process, there is a Turing machine or λ -calculus formal computation. In the present case, for every particular structure developed in the practice of mathematics, there is categorical characterization of it.

What are particular mathematical structures? Despite asserting e.g. that the natural numbers constitute a particular mathematical structure, I also say that the structure of the natural numbers (in the sense under discussion) is neither a mathematical object nor indeed an object of any sort. In that sense talk of “the structure of the natural numbers” is a *façon de parler*. In discussion Stewart Shapiro offered the suggestion that this construal of particular mathematical structures is a form of *if-then-ism*. Though if-then-ism is widely rejected as a philosophy of mathematics, I find this imputation helpful.

If-then-ism as a philosophy of mathematics originates with Bertrand Russell, in his first major book, *The Principles of Mathematics*, which opens with the declaration that “Pure Mathematics is the class of all propositions of the form ‘ p implies q ,’ where p and q are propositions containing one or more variables, the same in the two propositions, and neither p nor q contains any constants except logical constants” (Russell 1903, 3). Russell soon abandoned this view in favor of type-theory (initially simple, then ramified) as the means of understanding and overcoming the paradoxes. In 1967 Hilary Putnam considered “‘If-then-ism’ as a philosophy of mathematics” in his paper, “The Thesis that Mathematics is Logic” and declared a “wish to make a modest attempt to rehabilitate this point of view” (Putnam 1975, 20). He glosses Russell’s view by saying “What he meant was not, of course, that all well formed formulas in mathematics have horseshoe as the main connective! but that mathematicians are in the business of showing that *if* there is any structure which satisfies such-and-such axioms (e.g. the axioms of group theory), *then* that structure satisfies such-and-such further statements (some theorems of group theory or other).” I am happy with this formulation with the crucial caveat that Putnam’s example, in terms of axioms for a general structure, is badly chosen. Geoffrey Hellman, who rejects if-then-ism, considers exactly the sort of case with respect to which I wish to uphold a version of the if-then-ist viewpoint, namely when the antecedent characterize a particular structure, e.g. the natural numbers:

A categorical assumption to the effect that “ ω -sequences are possible” is indispensable and of fundamental importance. Without it, we would have a species of “if-then-ism,” i.e. a modal if-then-ism, and this would be open to quite decisive objections, analogous to those which can be brought against a naive, non-modal if-then interpretation. Consider the latter. Suppose it represents sentences A of arithmetic by means of a material conditional, say, of the form, $\wedge PA^2 \supset A$ or some refinement thereof. Suppose also that, *in fact*, there happen to be no actual ω -sequences, i.e. that the antecedent of these conditionals is false ... Then, automatically, the translation of every sentence A of the original language is counted as true, and the scheme must be rejected as wildly inaccurate. (Hellman 1989, 26)

I said above that I wish to uphold a *version* of the if-then-ist viewpoint, and this version is not *exactly* the position Hellman is dismissing. I agree that merely to represent true sentences of arithmetic as logical consequences of the second-order axioms of arithmetic is not sufficient. Something more is needed, but I disagree with Hellman over what the something more is.

For Hellman, what saves our account of truth in the structure of the natural numbers from vacuity is the “categorical modal existence assumption, ω -sequences are possible (1.4)” (1989, 19), about which he says “it is absolutely essential to affirm, categorically, an appropriate version of (1.4)” (27). He gives what purports to be a derivation of that “appropriate version of (1.4), which is (1.9)” (29–30). He writes down a condition (*) which expresses the existence of a discrete infinite sequence (29), about which he says, “there may be no reason to accept (*), but there is every reason to accept that, logically, it might be true” (30), and thereby takes $\Diamond(*)$ to be justified. The derivation he gives of (1.9) depends on $\Diamond(*)$, for which it has been claimed that we have “every reason to accept” but no reason has been given.

I find such modal claims empty. The something more that is needed to represent true sentences of arithmetic as logical consequences of the second-order axioms of arithmetic is not this, but the informal rigor by which we have come to understand these second-order axioms, and thereby to see that they are coherent. It is a development of mathematical understanding through informal rigor and not some further derivation that is needed. Hellman offers a clear account of the failings of Dedekind’s attempt to *prove* the existence of a simply infinite system by his (notorious) Theorem 66. What he is trying to do is give a different (of course, in his view, better) proof. But the point is, there can be no such *proof*. We must reflect on our conceptual understanding of a given particular mathematical structure as it has developed to see how it is that truths of e.g. arithmetic are those that hold in the structure of the natural numbers which we have succeeded in characterizing.

Having succeeded in this characterization, there is then the temptation to ask, “What are the mathematical objects that make up these structures?” Stewart Shapiro wants to answer this question by saying that they are positions in the structures we have succeeded in

characterizing. The difficulty with this derived notion of object is that it is a collapse back into Platonism, as Fraser MacBride has noted (2005, 582).

In response Shapiro claims (2008, 294) “Neither mathematics nor *ante rem* structuralism requires any more in the way of identification or individuation.” It seems to me that Shapiro cannot shake off this problem so easily. He cannot have objects without particularity. I agree with him entirely that there can be no requirement that they be identified or individuated by us.

The Euclidean plane brings out these issues. The structure of the Euclidean plane is categorical (as David Hilbert showed in his *Grundlagen der Geometrie*). But as Shapiro and others remark, given any two points, there is an automorphism that moves one to the other (translation or rotation—two different automorphisms) or swaps them (reflection). None of this is a problem for the notion of the structure of the Euclidean plane. But it is a problem for the idea that the individual points of the Euclidean plane constitute mathematical objects.

There are issues arising from the contrast between the method of construction vs. the method of postulation (I deliberately choose Russell’s terminology, while rejecting his vehement claims for superiority of construction over postulation).

Consider Hilbert’s axiomatization of the Euclidean plane vs. its construction by Cartesian coordinates.

Consider the theory of the complex plane as an algebraically closed and topologically complete field, vs. its construction as a set of ordered pair of reals with pointwise addition and multiplication defined to give it the structure of the complex numbers.

Consider the structure of the natural numbers given by second-order axiomatization by Dedekind, vs. the construction of the natural numbers within set theory, say as von Neumann ordinals.

Each of the constructions is well motivated but non-unique. The axiomatizations are also non-unique but provably equivalent.

In the context of the constructions there is no problem of automorphisms and identity of objects. The objects are what they are and not something else. E.g. $i = (1, 1)$, and $-i = (1, -1)$.

From Russell’s point of view (and Frege’s), the method of construction does indeed have many advantages over the method of postulation since Russell (and Frege) thought that logic constituted a

foundational bedrock on which these constructions could be erected. That viewpoint has been, in the form Russell (and Frege) held it, completely discredited. Second-order logic (pace Quine) is perfectly fine, indeed essential, but in combination with postulates. Nothing can be constructed from pure second-order logic by itself. What the method of construction achieves is relative interpretability of one theory within another, which is mathematically and conceptually significant, but not itself foundational.

Take the case of the Euclidean plane interpreted by Descartes as ordered pairs of real numbers (distances). It was mathematically hugely significant, as it allowed new methods to be applied to geometrical problems and became the basis for vast further developments (the calculus). However, it did not supplant Euclidean geometry if supplanting means that the supplanted theory ceases to be of any interest. Indeed, the very idea of supplanting Euclidean geometry in this way is conceptually incoherent since the only way in which it could be seen to supplant it is by a proof that all the truths of Euclidean geometry are interpretable within Cartesian geometry. But to establish such a result, we have to know exactly the content of Euclidean geometry. And that knowledge remains, indeed must be timelessly present, in the proof that Euclidean geometry can be interpreted within Cartesian geometry.

It seems to me that Euclidean geometry is particularly apposite to my claim that the structuralist understanding of mathematics must eschew any claim to establish the existence of mathematical objects, indeed is a basis on which to make it clear that there are no mathematical objects. This is because having characterized the Euclidean plane, there are no *particular* objects that have been characterized. The objects that there are, are purely the elements of that structure. There should be no temptation to suppose that we have access to any particular object.

The case of the natural numbers may misleadingly tempt object minded philosophers to think that there are particular objects, each individual natural number, since each number has a canonical term that refers to it, its numeral. This is not just an artifact of the countability of the natural numbers. Shapiro cites the case of the “cardinal-two structure” (2008, 12), where the structure has two objects but with no non-logical vocabulary there are no closed terms in the

language, so *a fortiori* no canonical terms that refer to the objects in that structure.

Shapiro's view is, in my view, dogmatism: "Still, I would insist, real numbers, members of the iterative hierarchy, points in Euclidean space, and places in finite cardinal structures and graphs, are legitimate mathematical objects" (2008, 8), and "My preference is to keep the notion of 'object' unequivocal between mathematical, scientific, and ordinary discourse" (9). The only support for these views that Shapiro cites there is that he is with Quine: "I follow Quine and take quantification as signaling ontology, rather than reference" (9).

The difference at issue here is reminiscent of the Quine-Carnap debate. From Carnap's viewpoint this talk of the existence of mathematical objects is confusion between internal and external questions. The claim is empty, mere metaphysics. This confusion is particularly clear in Shapiro's "faithfulness constraint" (2008, 6): "one desideratum of our enterprise is to provide an interpretation that takes as much as possible of what mathematicians say about their subject as literally true, understood at or near face value."

Note that despite the rejection of the existence of mathematical objects, the viewpoint (and it is a viewpoint rather than a theory) of this paper is *not* nominalist, for it accepts, indeed insists upon the reality of mathematical structures and the concepts in terms of which they are characterized (see Burgess and Rosen 1997). In "Mathematical Intuition and Objectivity" (Isaacson 1994), I distinguished between object Platonism and concept Platonism, and rejected the former while embracing the latter. It has since seemed to me that the label Platonism may not be apt, that the position I hold to is conceptual realism but not Platonism, since I see mathematics as an activity of mind and there is no sense (that I can see) of that conceptual reality existing in advance of, independently of, being conceived. Once conceived it has an (intersubjective) objectivity.

5 Full second-order quantification

Consider the difference between the second-order Axiom of Replacement (where ordered pairs of sets are defined in the usual way as $\langle x, y \rangle =_{df} \{\{x\}, \{x, y\}\}$:

$$\forall F \forall x (\text{In}(F) \supset \exists y \forall u (u \in y \equiv \exists v (v \in x \wedge \langle v, u \rangle \in F))) , \text{ where} \\ \text{In}(X) =_{df} \forall u \forall v \forall w (\langle u, v \rangle \in X \wedge \langle u, w \rangle \in X \supset v = w)$$

and the first-order Axiom schema of Replacement: For each formula ψ in the first-order language of set theory with free-variables u and v and possibly other free variables (parameters):

$$\forall x (\text{In}(\psi) \supset \exists y \forall u (u \in y \equiv \exists v (v \in x \wedge \psi(v, u)))) , \text{ where} \\ \text{In}(\psi) =_{df} \forall u \forall v \forall w (\psi(u, v) \wedge \psi(u, w) \supset v = w)$$

Kreisel holds that “the evidence of the first order schema derives from the second order axiom” (1967, 151).

The idea of a function from sets to sets does not carry with it the idea that it is definable in this or that language. We understand the idea of a real number, or a set of real numbers, in advance of grasping the notion of a definable real, or a definable set of reals. And we have no difficulty grasping the idea that there are reals (indeed most reals) that are undefinable in the usual language for expressing our understanding of real numbers.

A second-order language is always second-order relative to a first-order language. Insofar as we understand the first-order language, we understand the second-order language. But also, our understanding of what is expressed in the first-order language can be made precise by use of the corresponding second-order language. This analysis of the general situation is illustrated and tested by the most extreme case, that of set theory.

Set theory is extreme in two different ways. (1) Set theory is extremely general. It can encompass all structures. So there is reason to think that whatever we can say in general about structure cannot be made to apply to set theory. (2) The structure of the universe of sets is always partial and extensible.

Second-order quantification, i.e. quantification over collections (or properties) of objects, is something that occurs naturally, in both mathematics and ordinary language. Indeed, when mathematical logic was established by Frege, in 1879, the two existed together. Frege there used second-order logic to define and establish key properties of the ancestral of a relation. And with a certain amount of mathematical sophistication

but also naturally, the theory of the real numbers can be identified with second-order arithmetic, with the real numbers identified with sets of natural numbers. And there are examples in (fairly) ordinary language when second-order quantification is being used, i.e. needed to show perspicuously what is being said, as in this example by George Boolos: “There are some gunslingers each of whom has shot the right foot of at least one of the others.” Over the next fifty years attention focused on separating out first-order logic, culminating in the Gödel completeness theorem in 1930. Also, ways of formulating uses of second-order quantification in a prevailing first-order language were worked out, e.g. treating second-order quantification as first-order quantification over another sort, or translating everything into set theory.

In 1950 Leon Henkin showed how to extend the completeness theorem for first-order logic to second-order logic. The Henkin Completeness Theorem for second-order logic is far less decisive for second-order logic than Gödel’s completeness theorem for first-order logic since it does not deal with the *intended* meaning of second-order quantification, as ranging over *all* subsets of the first-order domain of objects. Rather what Henkin showed was the completeness of a deductive system for second-order logic with respect to *arbitrary* interpretations of the range of the second-order quantifiers. But the categoricity theorem for second-order arithmetic combined with Gödel’s incompleteness theorem in the form that true arithmetic is not recursively enumerable shows that there can be no complete system of *full* second-order logic. Indeed, the set of valid sentences in full second-order logic is highly undefinable—see Väänänen 2001, 517–518 for demonstration that it is not $\sum_n^m$ for any natural numbers m and n .

I have heard appeal to full second-order quantification derided as “a cheap trick.” The proper response is to distinguish clearly between deductions of theorems, for which formalized first-order logic is provably well-suited, and characterization of structures for which full second-order logical consequence is demonstrably well suited. (Hellman stresses this point [1989, 21] and cites Shapiro [1985] as also drawing this distinction.) Neither form of logic can serve the purpose to which the other is well-suited, and no treatment of logic can fulfill both. In that sense each is at cross-purposes with the other, but there is no incoherence in fulfilling both purposes by these different means. The very success of first-order logic for deduction unfits it for

characterization, and the very success of full second-order logic for characterization unfits it for deduction. The completeness theorem for first-order logic implies the compactness theorem and proofs of the completeness theorem establish the Löwenheim-Skolem Theorem. These two theorems show that first-order languages are not suited to characterizing structures. And as noted above, the success of characterizing the structure of the natural numbers by means of full second-order quantification establishes that there can be no complete system of deduction with respect to second-order validity.

There must be a tradeoff between competing desiderata on systems of logical deduction, and we simply must be clear what we are trying to do and on that basis make choices. Quine does not consider this duality of purpose and fiercely rejects second-order logic. For him, second-order logic is not logic at all, but “Set theory in sheep’s clothing” (1986, 66–68). The phrase imputes a touch of dishonesty on the part of one who cloaks second-order quantification in the sheep’s fleece of logic. On Quine’s view “the set theorists’s ontological excesses may sometimes escape public notice, we see, disguised as logic.” This attitude toward higher order logic embodies certain basic confusions. Quine standardly expresses the anxiety about set theory that one or another system with which we presently work may suddenly fail us, as did Frege’s by turning out to be inconsistent. This would come about through *ontological excess*, the adopting of principles that demand that inconsistent multiplicities (to use Cantor’s equanimous phrase in his letter to Rickard Dedekind) be unities, as happened to Frege’s *Grundgesetze*. Since Zermelo (1930) few, if any, set theorists have shared this anxiety, persuaded as they are of the cogency of the iterative conception of set (even if not persuaded by Zermelo’s categoricity theorems about that structure). Be that as it may, the point I wish to make here is that it seems to me quite misplaced to project these already quirky anxieties about set theory onto second-order logic. To do so is to fail to appreciate that there are crucial differences between second-order logic and set theory. The key point is the difference between multiplicities and unities (i.e. objects). This is where Frege’s Basic Law V goes wrong. As a principle about multiplicities it is indeed logically valid, a principle of logic.

In the discussion of Bernays 1967, Mostowski made the following point.

We need axioms which will characterize the notion of an arbitrary predicate. The solution of the continuum problem depends essentially on the choice of these axioms. But is the problem of their choice in essence not the same as the problem of finding suitable axioms for the notion of a set? Sets considered in mathematics are by their very nature second order objects. It does not seem to me that anything essential can be gained by the duplication of the problem and by stating it separately for sets considered as elements of the universe of an axiomatic theory and for sets considered as “extensional predicates.” (Bernays 1967, 114)

This is a particularly clear and trenchant expression of the view sloganized by Quine when he called second-order logic “set theory in sheep’s clothing.” It goes wrong already in its initial claim that “The solution of the continuum problem depends essentially on the choice of axioms which will characterize the notion of an arbitrary predicate,” which is refuted decisively by the point that the solution of problems undecided in the formalization of first-order arithmetic in PA but shown to be determinate by full second-order quantification does *not* depend on finding such axioms.

Second-order quantification is significant for philosophy of mathematics since it is the means by which mathematical structures may be characterized. But it is also significant for mathematics itself. It is the means by which the significant distinction can be made between the independence of Euclid’s fifth postulate from the other postulates of geometry and the independence of Cantor’s Continuum Hypothesis from the axioms of set theory. The independence of the fifth postulate reflects the fact, which can be expressed and established using second-order logic, that there are different geometries, in one of which the fifth postulate holds (is true), in others of which it is false. It makes no sense to ask whether the fifth postulate is really true or not. Whether it holds or not is a matter of which geometry we are in. The truth or falsity of the fifth postulate is not an open question, and is not something that can be overcome by finding a new axiom to settle it. By contrast, the independence of the continuum hypothesis does not establish the existence of a multiplicity of set theories. In a sense made precise and established by the use of second-order logic, there is only one set theory

of the continuum. It remains an open question whether in that set theory there is an infinite subset of the power set of the natural numbers that is not equinumerous with the whole power set.

6 The cumulative hierarchy as a particular structure

What is the metatheory within which Zermelo's quasi-categoricity theorem is proved? A possible answer is that it takes place within first-order ZF, and constitutes the study of a class of inner models for NBG namely those that are full. This was worked out by John Shepherdson in the second of his three papers on inner models (1952), in which he recast Zermelo's proof as a result about what he called "supercomplete" inner models of NBG. Akihiro Kanamori (1997) draws attention to this result when he sketches the Zermelo categoricity result in the first chapter of his book.¹⁵ Shepherdson compares his result to Zermelo's as follows:

Results essentially equivalent to theorems 3.12, 3.13, 3.14, 3.15 were obtained by Zermelo although in an insufficiently rigorous manner. He appeared to take no account of the relativity of set-theoretical concepts pointed out by Skolem, assuming that such concepts as sum set, power set, cardinal number etc, had an absolute significance. Thus what he considered to be the only possible kinds of inner model are what we should describe as super-complete models. Apart from these logical shortcomings the proofs he gave of his results are in outline the same as those given here. (Shepherdson 1952, 227)

The inner model perspective also provides a basis for understanding the status of the axiom of foundation. The axiom of foundation is immediately seen as true in the intended structure of the cumulative hierarchy of sets. Even so, it is also known that there are non-wellfounded sets, i.e. structures in which the other axioms of ZF hold in which the Axiom of Foundation is false—cf. Peter Aczel (1988). What we can

¹⁵ Ferenc Csaba cited Shepherdson's result in his commentary on my Budapest lecture.

show in ZF without Axiom of Foundation is that in the sub-domain of the well-founded sets of any domain all the axioms of ZF, including Foundation, hold. This result establishes the relative consistency of the Axiom of Foundation with the other axioms of ZF, i.e. if we denote ZF minus the Axiom of Foundation by ZF^- , then

$$ZF^- \vdash (Con\ ZF^- \rightarrow Con\ (ZF^- + AF))$$

By contrast, for $I =$ “There exists an inaccessible cardinal”

$$ZF \not\vdash (Con\ ZF \rightarrow Con\ (ZF + I))$$

This latter result follows from the proof (in ZF) that if κ is an inaccessible ordinal, then V_κ is a model of ZF, so $(ZF + I) \vdash Con\ ZF$. If $ZF \vdash (Con\ ZF \rightarrow Con\ (ZF + I))$, then *a fortiori* $ZF + I \vdash (Con\ ZF \rightarrow Con\ (ZF + I))$, in which case, by *modus ponens*, $ZF + I \vdash Con\ (ZF + I)$, which by Gödel’s Second Incompleteness Theorem is impossible if $ZF + I$ is consistent.

In the sense of my papers on arithmetical truth (Isaacson 1987, 1992), this result is indicative of the situation that the notion (and existence) of an inaccessible cardinal is a “hidden higher-order concept” with respect to the notion of set.

The ZF axioms for pure sets with second-order Replacement determine everything about the cumulative hierarchy of sets except its height. Rather than a weakness, i.e. only having quasi-categoricity vs. strict categoricity, this is intrinsic to the notion of set, namely as reflecting its indefinite extensibility. Zermelo insists on this point at the end of his paper:

Scientific reactionaries and anti-mathematicians have so eagerly and lovingly appealed to the “ultrafinite antinomies” in their struggle against set theory. But these are only apparent “contradictions,” and depend solely on confusing *set theory itself*, which is not categorically determined by its axioms, with individual *models* representing it. What appears as an “ultrafinite non- or super-set” in one model is, in the succeeding model, a perfectly good, valid set with both a cardinal number and an ordinal type, and is itself a foundation stone for the construction of a new domain.

To the unbounded series of Cantor ordinals there corresponds a similarly unbounded double-series of essentially different set-theoretic models, in each of which the whole classical theory is expressed. [my emphasis] The polar opposite tendencies of the thinking spirit, the idea of creative advance and that of collection and completion, ideas which also lie behind the Kantian “antinomies,” find their symbolic representation and their symbolic reconciliation in the transfinite number series based on the concept of well-ordering. This series reaches no true completion in its unrestricted advance, but possesses only relative stopping-points, just those “boundary numbers” which separate the higher model types from the lower. Thus the set-theoretic “antinomies,” when correctly understood, do not lead to a cramping and mutilation of mathematical science, but rather to an, as yet, unsurveyable unfolding and enriching of that science. (Zermelo 1930, 1233)

Zermelo’s insight is expressed exactly by Tarski’s Axiom of Inaccessible Cardinals (which, shorn of combinatorial aspects of Tarski’s formulation designed to imply the Axiom of Choice)¹⁶ says:

$$\forall \alpha \exists n (Inac(n) \wedge \alpha < n)$$

The following exercise given by Frank Drake (1974, 111) provides the basis for an account of this axiom:

Suppose that the strongly inaccessible cardinals have been enumerated in order as θ_α for ordinal α . Then a cardinal κ is hyperinaccessible if $\theta_\kappa = \kappa$, i.e. if κ is strongly inaccessible and there are κ strongly inaccessible cardinals below κ .

Drake’s exercise is to show that if κ is the first hyperinaccessible cardinal, then V_κ is a model of ZFC + the axiom of inaccessibles + “there are no hyperinaccessible cardinals.”

For ZFC² the converse holds.

On the one hand this gives a certain conceptual clarity to the Zermelo insight, i.e. a finite expression to its content. This situation corresponds exactly to Zermelo’s original insight that $V_\kappa \models \text{ZFC}^2 +$ “there is no strongly inaccessible cardinal” if κ is the smallest strongly

¹⁶ I am grateful to Philip Welch for sorting out this feature of Tarski’s axiom for me.

inaccessible cardinal, which latter condition is equivalent to the condition that κ is the smallest *regular* fixed point of the normal function $\alpha \mapsto \aleph_\alpha$.

Zermelo's insight that is now under discussion was that this result is not stable, which one can say is obvious since we are modeling "there is no strongly inaccessible cardinal" with a construction based on a strongly inaccessible cardinal. This instability pushes us to iterate this understanding, i.e. Zermelo's insight.

The corresponding iteration of Zermelo's insight leads us to an Axiom of Hyperinaccessibles, i.e. a sequence of hyperinaccessibles indexed by all the ordinals,

This in turn leads to an Axiom of Hyper-hyperinaccessibles. And so on.

As Frank Drake notes (1974, 115), there is a single axiom which implies all such axioms: Axiom F. Every normal function has a regular fixed point.

This axiom implies every axiom of the sort that arise from iterations of Zermelo's insight.

The indefinite extensibility of the set concept shows that there is no quantification over "absolutely everything." Against this point it might be argued that the very attempt to express the indefinite extensibility claim, e.g. "For every domain of sets that satisfy the axioms of ZF^2 , there is a domain in which the original domain is a set," or more precisely, for every inaccessible ordinal κ , $V_\kappa \models ZF^2$ requires that we understand the quantifier "for every inaccessible ordinal κ " as a universal quantifier ranging over all sets. This is a natural first idea, but is untenable.¹⁷ The way the indefinite extensibility claim has to be construed is as quantification over *understandings* of the concept of set, and with each understanding is associated a domain of sets. Whatever understanding we have of the set concept, we can, on that basis, arrive at an understanding that extends the domain of sets in such a way that the domain of the understanding we first arrived at is subsumed as a set. This construal does not tempt us to think that there is, and must be, a

¹⁷ There is a curious dynamic here. The advocate of a fixed domain of all sets considers that the indefinite extensibility of sets requires a fixed domain of all sets, and the opponent of the idea of a fixed domain of all sets considers that the indefinite extensibility of sets establishes that there is no fixed domain of all sets.

domain of all sets that there are, as clearly there is no fixed domain of all understandings of set theory. An understanding exists when it has been understood.

If the quasi-categoricity result is formulated in terms of inner models, as established by Shepherdson, we have a different basis on which to reject the argument that the indefinite extensibility of sets requires that there be a fixed domain of all sets. The point is simply that on this basis the quantifier “for every inaccessible ordinal κ ” in the result that for every inaccessible ordinal κ , $V_\kappa \models \text{ZF}^2$ is relative to whatever understanding we have of the extent of the universe.

How is it that for set theory quasi-categoricity suffices to establish a particular structure whereas in the case of other particular structures categoricity is required? My answer is that what is undecided in virtue of this degree of non-categoricity is genuinely undecided, in the same way that the fifth postulate of Euclid’s geometry is genuinely undecided by the other axioms. This includes GCH (or some version that survives the refutation of CH), and the existence of large cardinals (among which the first uncountable inaccessible is teensy—to use George Boolos’s term [1998, 120]).

7 The status of Cantor’s continuum problem

Besides their implications for philosophy of mathematics, these results also have mathematical significance, by making it clear that establishing the size of the continuum is a mathematical problem with a determinate answer which set theorists of sufficient genius can hope to solve. This view, though far from universally acknowledged, is not news. Gödel, having anticipated the independence of CH, firmly held to its being a genuine problem in the aftermath of Cohen’s results and continued to attempt to solve it (see Gödel 1995). In the present day Hugh Woodin, one of the world’s leading set-theorists, is pursuing a program whose outcome, if ultimately successful will be to show that $2^{\aleph_0} = \aleph_2$ (see Woodin 2001, 2005).

The nature of the continuum problem is illuminated by comparing it with other problems not solvable by specified conceptual resources.

(1) Borel Determinacy. Harvey Friedman (1971) showed that this cannot be settled in Zermelo’s original system, Z, formulated in

“Untersuchungen über die Grundlagen” (1908). D.A. Martin (1975) then proved it using the Axiom of Replacement, i.e. in ZFC. This result should be compared with Projective Determinacy, which is independent from ZFC, but is proved from the existence of infinitely many Woodin cardinals (Martin and Steel 1988).

(2) Kreisel incompleteness theorem. Kreisel’s construction yields Δ_2 -sentences that are independent of the system for which they are constructed, for example PA (see Kreisel 1950 and 1968, 382; see also Smorynski 1977, 860–864). The truth-value of a few contrived examples of these can be established, but in general they are of unknown truth-value. See Manevitz and Stavi (1980), which opens with the following remark:

Determining the truth-value of self-referential sentences is an interesting and often tricky problem. The Gödel sentence, asserting its own unprovability in P (Peano Arithmetic) is clearly true in N (the standard model of P), and Löb showed that a sentence asserting its own provability in P is also true in N ... The problem is more difficult, and still unsolved, for sentences of the kind constructed by Kreisel (1950).

We are confident that these Kreisel sentences are true or false, even when we know that they cannot be decided in PA and have no particular idea of principles by which they can be decided.

There is also a likely disanalogy between the determinateness of the continuum problem and the two cases just cited. The Axiom of Replacement by which Zermelo’s original axiomatization of set theory had to be extended if Borel Determinacy was to be proved is formally simple and conceptually easy to grasp. In the case of Kreisel’s Δ_2^0 -sentences of arithmetic undecidable in PA, say, the difficulty is not in finding a new axiom to settle them but in finding the means by which to handle the combinatorial complexity involved in deciding them. To judge by the understanding of long chains of set-theoretic results required to grasp Woodin’s Ω -conjecture, the likelihood is that grasping the principles, let alone to be able to prove from them that $2^{\aleph_0} = \aleph_2$, as Woodin expects, will require a very great deal more expertise in set theory than is needed for any hitherto proposed axioms. This is in contrast with the way in which set theory has hitherto, since

Zermelo's original axiomatization, been extended, namely by axioms of infinity. (The Axiom of Replacement has the effect, even though not the form, of an axiom of infinity.) The existence of an inaccessible cardinal is transparently motivated by our understanding of the cumulative hierarchy of sets, and Mahlo cardinals follow on naturally from that. Beyond Mahlo, new ideas are needed but still the nature of the extension of set theory by such axioms is clear. In 1946 Gödel was of the view that every statement that could be proved in an extension of set theory is provable by an axiom of infinity:

In set theory, e.g. the successive extensions can most conveniently be represented by stronger and stronger axioms of infinity ... It is not impossible that for such a concept of demonstrability some completeness theorem would hold which would say that every proposition expressible in set theory is decidable from the present axioms plus some true assertion about the largeness of the universe of all sets. (1965, 151)

After Cohen's result, and developments from it, it became clear essentially that the continuum problem cannot be settled by any axiom of infinity (see Martin 1976, 86). Thus whatever principle settles the continuum problem must be of a new kind that goes beyond our understanding that the universe of sets constitutes an indefinitely extensible cumulative hierarchy. One may conjecture that this situation reflects the fact that the continuum problem is independent of systems that arise from characterization of the concept of set.

Crispin Wright, commenting on my Munich lecture, raised an issue about my form of argument for determinateness of problems in the theory of natural numbers and in set theory that shows there is a realist assumption in my viewpoint that from a constructivist, anti-realist, point of view, is question-begging." "You need to assume that once a structure is determinate so is the notion of truth for all sentences in the structure, in order to argue that the Kreisel Δ_2 -sentences must likewise be determinate, and the corresponding assumption in the case of the set theoretic universe." Wright's point is that someone who is worried about indeterminacy will not just find it obvious that where a structure is determinate we can take it that, as it were, the sentences will take care of themselves, especially with respect

to quantification. This is an old and central debate in the establishment of intuitionistic mathematics. Suppose we consider that the natural numbers are generated by successive acts of intuition. Why is not quantification over that domain determined? It is even the case that Dedekind's axioms for the natural numbers hold for the natural numbers, so understood. Even the notion of logical consequence is cognate, i.e. $\Gamma \models \varphi$ if in every structure in which all the formulas of Γ come out true, φ comes out true. I then want to say that for a set of sentences Γ which comes out true in only one structure (to within isomorphism), every sentence φ in the language of Γ , either $\Gamma \models \varphi$ or $\Gamma \models \neg\varphi$. But in saying this I am relying on the law of excluded middle, i.e. that for every formula φ in the language of a particular structure, either φ is true in that structure or $\neg\varphi$ is true in that structure. (For my response see section 2.)

The reality of mathematics and the connection of that issue with the phenomenon of incompleteness is a live issue today, both among mathematicians and philosophers, as shown in the following quotations. First, from Timothy Gowers, a Fields Medalist at Cambridge (2006, 193):

I can date my own conversion from an unthinking childhood Platonism from the moment when I learnt that the continuum hypothesis was independent of the other axioms of set theory. If an apparently concrete statement as that can neither be proved nor disproved, then what grounds can there be for saying that it is true or that it is false?

And a philosopher, Hartry Field (2000, 25):

A possible view is that even though our axioms don't settle the matter (nor do any extension of the axioms that we find at all "evident"), still our set theoretic concepts are perfectly precise, so that there is an objectively correct answer that we will probably never know. But an alternative view, which I find far more plausible, is that our set-theoretic concepts are indeterminate: we can adopt any one of the above claims about the size of the continuum we choose, without danger of error, for our prior set-theoretic concepts aren't determinate enough to rule the answer out.

Serge Lang, a notable mathematician:

when you speak of “sets,” you don’t know what you are talking about. The ambiguity lies in the intuitive notion you have of a set. Everybody has some intuition of sets: a set is a ... bunch of things. [Laughter.] To say a bunch of things, it’s OK if you speak of all the real numbers; it’s OK if you speak of all the rational numbers; it’s OK if you speak of all the points on a curve; but if you speak of all sets simultaneously, of all the sets contained in the real numbers, then it’s not OK, it does not work any more. That’s what Paul Cohen’s answer means: our notion of set is too vague for the continuum hypothesis to have a positive or negative answer. (Lang 1985, 54; ellipsis and brackets in the original)

Solomon Feferman (1999, 109): “I am convinced that the Continuum Hypothesis is an inherently vague problem that *no* new axiom will settle in a convincingly definite way.”

Also from Solomon Feferman (2000, 405): “My own view—as is widely known—is that the Continuum Hypothesis is what I have called an ‘inherently vague’ statement, and that the continuum itself, or equivalently the power set of the natural numbers, is *not* a definite mathematical object.”

An article by Paul Cohen in collaboration with Reuben Hersh declares: “We propose in this article to use the oft-told tale of non-Euclidean geometry to illuminate the now unfolding story of non-standard set theories” (Cohen and Hersh 1967, 104). (‘Non-standard’ is used here as being analogous with non-Euclidean, rather than with the meaning of non-standard in the case of non-standard models of arithmetic.)

Kreisel rejects this claimed analogy between set theory and geometry. In “Informal Rigor and Completeness Proofs” (1967) he points out the precise reason why these two independence results are different in kind, namely that the parallel postulate is not determined by a second-order formulation of the principles of geometry, and remains independent in that formulation, whereas the continuum hypothesis is determined in the second-order characterization. It seems to me that this gives a completely perspicuous account of the essential difference in the conceptual character of the set-theoretic result and the result in geometry. In the one case we are concerned with genuine alternatives,

differences that require choices. There are different structures in geometry with which we may be concerned, while in the case of set theory there is a single structure, which we understand well enough to be able to characterize it precisely, but for which we then see that certain questions which we are able to formulate in the language of that structure are nonetheless not determined by the principles which result naturally from our successful attempt to characterize that structure. At the same time Kreisel is not claiming that the logical consequences of ZFC² are complete. He notes (in his review of Weston 1977) that ZF² gives no basis for taking it that the GCH is determinate (it might hold up to the first inaccessible and fail thereafter).

Some responses to the second-order categoricity result have been skeptical. But such skepticism is misplaced if it is based on the supposition that it is meant to have anything to do with being able to determine the truth or falsity of the continuum hypothesis. Rather it allows us to see that we understand the situation, so that having identified the question as Cantor did and analyzed the concept of set in the way that Zermelo did, we are certain that we have a well-defined question. The independence results show that solving this problem requires new axioms.

Kalmár, in the meeting at which Kreisel presented “Informal Rigor and Completeness Proofs,” declared:

I have the impression that for Kreisel second order methods are as useful and as correct as first order ones. I do not agree with him ... I think second order categoricity results are *deceiving*. They serve only to puzzle ordinary mathematicians who do not know enough logic to distinguish between first order and second order methods ... I do not think second order categoricity theorems can serve any sound purpose. (Kalmár 1967, 104)

Andrzej Mostowski noted that the results by Cohen and others concerning the continuum hypothesis “show that practically every hypothesis concerning powers of regular cardinals is compatible with the axioms of Zermelo-Fraenkel and declared that, “Such results show that axiomatic set theory is hopelessly incomplete” (1967, 93). He went on to say that the independence results show that there is no single set theory within which the continuum hypothesis is a determinate

problem, and cites the split between different notions of space in Euclidean and non-Euclidean geometries as showing the situation set theory has now been shown to be in:

Probably we shall have in the future essentially different intuitive notions of sets just as we have different notions of space, and will base our discussions of sets on axioms which correspond to the kind of sets which we want to study. (94)

It is quite true that set theorists base their discussions on axioms which correspond to the kind of sets they want to study, for example the constructible universe L or the forcing model of Solovay in which all sets of reals are measurable. But this is with an awareness that these are models of set theory and not set theory itself.

The axiom $V=L$ proves the continuum hypothesis but no one takes this as establishing that the continuum hypothesis is true. The constructible universe is seen to be a *restriction* of the whole universe.

If the set notion were relative in this way then indeed study of the continuum problem should and would have come to an end as Gödel's and Cohen's results came to be understood, in the same way that attempts to prove the fifth postulate of Euclidean geometry came to an end after the work of Bolyai and Lobachevsky was known and understood. This has not happened.

Mostowski held that set theory is *not* "a clear and well-understood branch of science" (82) (for which reason Mostowski considered that the reduction of mathematics to set theory does not "provide us with a satisfactory basis for mathematics," a viewpoint that had been advanced on more general, i.e. logical grounds thirty years earlier by Skolem, who rejected set theory as a foundation of mathematics because of the relativity of its notions as revealed by the Löwenheim-Skolem theorem, i.e. Skolem's so-called paradox).

Woodin has got on with the project of settling the continuum problem without (so far as I am aware) having given general arguments to establish that it is a genuine problem. Gödel famously gave such arguments, in his paper "What is Cantor's Continuum Problem?" (1947). Gödel's arguments are inductive and heuristic, showing how the continuum problem interacts with existing developments within mathematics, but also conceptual.

I conclude by confronting four arguments that have been raised against the viewpoint developed in this paper.

(A) The claimed determinateness of the continuum problem on the basis of the quasi-categoricity of second-order set theory has been challenged on the grounds that Gödel did not use this argument, from which fact, so it is claimed, we can infer that he rejected it, since if he had considered it to be cogent he would have used it in “What Is Cantor’s Continuum Problem?” (1947) as being more decisive than the indirect arguments he marshaled in that paper. And if indeed, as this argument shows, Gödel rejected this use of quasi-categoricity of set theory, then those who now reject it are in good company.¹⁸ The point is *ad hominem* and can be answered in kind, but there are also interpretative considerations of “What Is Cantor’s Continuum Problem?” that are an answer to it.

Zermelo’s conception of logic was infinitary and, infamously, he did not understand Gödel’s incompleteness theorems (1931) when the paper was published, nor from hearing a lecture by Gödel at a meeting of the Deutsche Mathematiker-Vereinigung held in Bad Elster 15 September 1931, at which they both gave lectures. On 21 September 1931 Zermelo wrote to Gödel enclosing “my *Fundamenta* paper” almost certainly (1930), and declaring “I would be pleased if I might count you among the few who have least tried to take up the ideas and methods developed there and make them fruitful for their own research.” He goes on to inform Gödel that “while I was engaged in preparing a short abstract of my Elster lecture, in the course of which I had also to refer to yours, I came subsequently to the clear realization that your proof of existence of undecidable propositions exhibits an essential *gap*” (Gödel and Zermelo 1931, 421), which he went on to expound. Gödel replied on 12 October 1931 with a patient explanation, running to three printed pages, of his proof of incompleteness.

That Gödel should want particularly to make sure that Zermelo understood and accepted his result is easily inferred from the fact that the “related systems” to which he refers in the title of “Über formal untentscheidbare Sätze der *Principia mathematica* und verwandter Systeme I” is “the Zermelo-Fraenkel axiom system of set theory (further developed by J. von Neumann)” (Gödel 1931, 145) (which he

¹⁸ Something like this argument was put to me by Penelope Maddy in discussion.

characterizes as being along with *Principia Mathematica* “so comprehensive that in them all methods of proof today used in mathematics are formalized, that is, reduced to a few axioms and rules of inference”) and “as can easily be verified, included among the systems satisfying the assumptions 1 and 2 are the Zermelo-Fraenkel and the von Neumann axiom systems of set theory” (181). (Note, incidentally, that he also remarks that the copy of his *Fundamenta* paper Zermelo was sending to him had not arrived, but “anyway, I had already read your paper soon after its appearance and at that time various thoughts occurred to me which, if they interest you, I will be glad to impart to you next time” [Gödel and Zermelo 1931, 429].) Zermelo replied on 29 October 1931, thanking Gödel for his “friendly letter, from which I can now infer, better than from your paper and your lecture, what you really mean” (Gödel and Zermelo 1931, 431). But he still insists that Gödel’s argument depends on a “finitistic restriction” and insists that without this restriction “you obtain an *uncountable* system of possible statements, among which only a *countable* subset are ‘provable’ and there must certainly be ‘undecidable’ statements.” Gödel cannot have been happy at this obdurate incomprehension of his result but worse seems to be Zermelo’s apparent rudeness in ignoring Gödel’s offer to send his thought on Zermelo’s *Fundamenta* paper, and it is the latter that we must especially regret, since the upshot is that Gödel did not communicate his thoughts on Zermelo’s categoricity result for set theory in a letter which it is likely (given what else has survived) we would now have in Gödel’s published correspondence.

Gregory Moore notes (1980, 128) that “When Zermelo published Cantor’s collected works in 1932 he regretted that Cantor the mathematician and Frege the logician had so little understood and appreciated each other (Zermelo 1932, 442). Yet similar misunderstandings marred the relationship between Zermelo and Gödel.” Given that Zermelo took it upon himself to send (or at least planned to send) a copy of his *Fundamenta* paper to Gödel, and his stated keenness to enlist Gödel as someone who would take up the “ideas and methods developed there,” it is not possible without more information to explain his failure to accept Gödel’s offer to send his thoughts on that paper.

I put Maddy’s point to Kreisel, to which he replied that in his conversations with Gödel (in the 1950s), Gödel had readily agreed with his observation that the continuum is determined by the second-order

axioms, and this was evidently a thought Gödel had had himself.¹⁹ Why then did Gödel not espouse this argument himself? There are several factors which, it seems to me, suggest an answer. One is the unsatisfactory correspondence with Zermelo in conjunction with Zermelo's key role in propounding this argument. I do not mean that Gödel was miffed by Zermelo's incomprehension and seeming rudeness. The seeming rudeness must be beside the point, but the incomprehension of Gödel's incompleteness theorems is not.

There was an apparent link between Zermelo's espousal of full second-order consequence as a basis for his categoricity theorems and his inability to understand Gödel's incompleteness theorem. How could he cite the result to an audience who he would expect to understand the metamathematical precision of the sort by which he was able to obtain his incompleteness theorems? Well, it might be said, Gödel could have explained Zermelo's result in his paper on Cantor's continuum problem (Gödel 1947) in the satisfactory terms he understood it himself. Here, however, Gödel's noted cautiousness suggests otherwise (see Feferman 1984). He would have had to advocate an understanding of logic that was already in 1931 and certainly by 1947 almost entirely alien. It can be advocated in conjunction with the received understanding of logic by distinguishing sharply between the theorem-proving function of first-order logic and structure-characterizing function of full second-order quantification. But this argument would not have fitted within an established philosophical tradition. Some might find this situation no impediment to pressing their point. But not Gödel. Compare his ultimate refusal to publish a paper on Carnap's conception of mathematics as syntax of language (for the *Library of Living Philosophers* volume on Carnap), on which he worked for six years, at the end of which time he wrote to the editor, Paul Arthur Schilpp to say that he would not submit his paper for publication (Gödel 1959).

The fact is that I have completed several different versions but none of them satisfies me. It is easy to allege very weighty and

¹⁹ Kreisel added that Gödel had not noted that the second-order axioms do not determine the generalized continuum hypothesis, which Kreisel pointed out to him, and which he found striking.

striking arguments in favor of my views, but a complete elucidation of the situation turned out to be more difficult than I had anticipated, doubtless in consequence of the fact that the subject matter is closely related to, and in part identical with, one of the basic problems of philosophy, namely the question of the objective reality of concepts and relations. On the other hand, because of widely held prejudices, it may do more harm than good to publish half done work.

Warren Goldfarb (1995, 324) comments on this passage, “Here Gödel is displaying his characteristic caution (see Feferman 1984), as well as his overestimation of the extent to which positivist dogma—what he means by ‘widely held prejudices’—remained orthodoxy by 1959.” Further to this theme Goldfarb cites Gregory Moore’s Introductory note to Gödel’s “What Is Cantor’s Continuum Problem?” (Moore 1990, 166), where Moore reports that when Benacerraf and Putnam asked permission to reprint his paper on Cantor’s continuum problem, “at first, Gödel hesitated to grant permission, fearing that the introduction to their book would subject his article to positivistic attacks.”

However, Gödel’s fears about “positivistic attacks” might not have been unfounded had he attempted to deploy Zermelo’s proof of quasi-categoricity of ZFC by use of full second-order logical consequence, as the published responses to Kreisel’s deployment of that argument (1967) make evident.

The final crucial point, however, is interpretative and conceptual rather than personal. Key passages in “What Is Cantor’s Continuum Problem?” express Gödel’s strong conviction in the reality of set theory. These passages have been much argued over and various commentators have held them to be an ill-founded espousal of mystical Platonism. For example:

For if the meanings of the primitive terms of set theory as explained on page 262 and in footnote 14 are accepted as sound, it follows that the set-theoretical concepts and theorems describe some well-determined reality, in which Cantor’s conjecture must be either true or false. Hence its undecidability from the axioms being assumed today can only mean that these

axioms do not contain a complete description of that reality.
(Gödel 1947, 260)

We make much better sense of Gödel's confidence in the "well-determined reality" described by set theory, especially when we see from other passages that he takes this reality to be constituted by the cumulative hierarchy if we suppose him to have in mind as the basis for this confidence a clear and specific argument rather than mystical confidence in an otherwise unexplained intuition.

At several points Gödel's language is very close to that of Zermelo (1930). For example:

It follows at once from this explanation of the term 'set' that a set of all sets or other sets of a similar extension cannot exist, since every set obtained in this way immediately gives rise to further applications of the operation "set of" and therefore to the existence of larger sets. (259, fn. 15)

and also

The simplest of these strong "axioms of infinity" asserts the existence of inaccessible numbers (in the weaker or stronger sense) $> \aleph_0$. The latter axiom, roughly speaking, means nothing else but that the totality of sets obtainable by use of the procedures of formation of sets expressed in the other axioms forms again a set (and, therefore, a new basis for further applications of these procedures). (260)

I do not suppose that Gödel was specifically following Zermelo in thinking this way. He was clearly capable of thinking these things through for himself. We know from Kreisel's account that Gödel did think in these terms, and we have seen *ad hominem* reasons why he would not have stressed this aspect of his thinking. I conclude that far from rejecting the conception of the reality of the subject matter of set theory developed in this paper, Gödel should be seen as having used it as the basis for his thinking about the nature of sets.

(B) Categoricity using full second-order logic to ensure the definiteness of the power set operation over ω and so the definiteness of

the continuum problem, is circular since the full second-order quantifier ranges over the powerset of whatever the domain is.

Answer to (B): We do not rely for these arguments on “definiteness” of the range of the second-order quantifiers. They *express* understanding. The notion of sub-multiplicity is what is required, which is *not* the notion of subset. Still less is it in any way dependent on the existence of powerset. We understand perfectly well the idea that the multiplicity of all limit ordinals is a sub-multiplicity of the multiplicity of all ordinals. This example also shows the relativity of multiplicities to the ambient domain. The multiplicity of all ordinals is always relative to a domain, a V_κ for κ inaccessible. Take as an example of the facility with which we operate with these notions Lemma 6.9 of Jech (2003, 67): If E is a well founded relation on a class P , then every nonempty class $C \subseteq P$ has an E -minimal element. This point requires careful formulation. At that point in Jech’s book, and standardly, classes are taken as given by formulas. They are multiplicities that may not be unities (sets). But they are defined by a formula in the first-order language of set theory. Thus in a sense we are merely talking about a *predicative* extension of first-order set theory. This situation is spelled out in detail (not in exactly these terms) by Levy (1979, sections 3 and 4 of chapter 1 and the Appendix), proving that such an extension is conservative over first-order ZF. This is, for my purposes in this paper, too restricted. We need a *full* second-order extension of the language of first-order set theory. From earlier discussion clearly this is a semantic and not a syntactic (formal) notion. The point is that we can make this extension by *dropping* the interpretation on this second-order quantification that every class is given by a formula in the language of set theory. The argument for the Lemma remains fully cogent without that supposition. All that it relies on is that C is non-empty and $C \subseteq P$. The proof goes through just as before, and is fully understandable in these unrestricted terms.

A domain of quantification is a multiplicity which may or may not be a unity. In particular V is an inconsistent multiplicity, yet we formulate Cantor’s theorem, as “for every set, there is no bijection between it and its power set.” All ordinals constitute an inconsistent multiplicity, and “every ordinal has a successor.”

(C) Even if (B) is answered, categoricity only proves uniqueness, *not existence*.

Answer to (C): The challenge is met by noting that categoricity articulates our understanding of structure and establishes that we understand the structures for which these results are established.

(D) The reaction of working mathematicians to the existence of undecidable sentences of arithmetic, such as the Gödel sentence, is almost never that such sentences thereby do not have a truth-value. Whereas this reaction to the independence of CH among mathematicians is common and even among some specialists in set theory. Is it philosophical obtuseness that prevents them grasping the similar nature of sentences undecidable in systems of arithmetic and sentences undecidable in systems of set theory, or is there is a deeper reason why the undecidability result about arithmetic almost never provokes this kind of reaction whereas the ones about set theory do?²⁰

My reply is that undecidability of sentences in the language of arithmetic by particular axioms systems does not provoke the same reaction as corresponding results for set theory because no one thinks they do not understand the notion of natural number. Whereas it is much less common for anyone to think, unreflectively, that they understand the notion of a pure set. There is then the question whether this is an accidental artefact of the history of set theory and logic, by which the distraction of Russell's paradox was given seeming relevance by Frege's ill conceived Basic Law V, or whether it reflects something intrinsic to the notion of set that makes it different from number theory in a way that justifies these differing responses. There are differences between the notion of set and the notion of number that may explain but they do not justify these differing reactions.

Arithmetic is an ancient subject, set theory relatively recent. Every minimally educated person knows some arithmetic, and from early childhood education. Basic arithmetical facts (in the standard model) are constantly applied in everyday life by everyone. Very differently, results in set theory are, if encountered at all, only encountered in late adolescence.²¹ Few people ever know anything about pure set theory.

²⁰ Something like this argument was raised in discussion by John Burgess.

²¹ The remoteness of set theory from immediate experience is exemplified in a curious paper by George Boolos (1998), in which he expresses serious doubts about the intelligibility of the full infinitary concept of (pure) set. What is curious about this paper is that someone as logically knowledgeable and sophisti-

Set theory is a powerful theory in which arithmetic can be interpreted. The only part of set theory that can be interpreted in arithmetic is the theory of hereditarily finite sets. Insofar as infinity is understood at all, the arithmetic of the natural numbers is in some sense fully understood. The indefinite extensibility of the hierarchy of sets means that in some sense the set theory of the cumulative hierarchy can never be fully understood. Even those who know something about set theory often, knowing of Russell's paradox, think dimly that there is something murky and confused about it.

For all this, the full cumulative hierarchy, V , is the intended model of set theory in the way that N is *the* intended model of arithmetic. What the student of set theory first encounters is the standard model, the cumulative hierarchy (if taught at all well; see, for example, Enderton 1977). In the initial stages there is no talk of models, only of sets. Later, in a more advanced course, the student encounters inner model constructions and in particular the constructible universe L . Still later come forcing models. Working set theorists are almost always, one way or another, concerned with non-standard models of ZFC with large cardinal axioms, in contrast to number theorists, who are rarely concerned with non-standard models of arithmetic and whose results can almost always be formalized in PA.

None of these differences give any good reason to infer from the fact that CH is undecided in first-order set theory to the conclusion that there is no fact of the matter as to the truth or falsity of CH. We see this clearly by considering the case of Borel Determinacy. Application to this case of the argument under consideration would lead from undecidability of Borel Determinacy in Z to the mistaken conclusion that there is no mathematical fact of the matter as to the Determinacy of the Borel sets. From the encompassing perspective of Zermelo-Fraenkel set theory we see that any such conclusion would have been mistaken.

cated as George Boolos should choose to express such a naive viewpoint—for this aspect of the paper see discussion of it by Paul Benacerraf (1999). But taking this paper at face value, it exemplifies the remoteness of set theory from immediate experience.

8 Conclusion

Recognizing the subject matter of mathematical theories as structures in the way expounded in this paper provides a perspective on the nature of mathematics that allows us to understand what mathematics is about and how what it is about constitutes a reality.

Acknowledgements and dedication. I am grateful to Zsolt Novák and András Simonyi for their invitation to participate in the conference for which this volume is the proceedings, and tremendously indebted to them for their thoughtful comments on an earlier draft of this paper and for their highly perceptive detailed editing of the penultimate draft, and for their kind patience and encouragement while I have worked on this paper. I have benefited from other invitations to speak in conferences and seminars where I have presented several of these ideas. The first was from Matthias Schirn, to whom I am greatly indebted for inviting me to speak in his conference on The Philosophy of Mathematics Today. Other invitations have come from Petr Hajek, Fabrice Pataut, Marco Panza, Paolo Mancosu, Michael Potter, Marcus Rossberg, James Ladyman, and Francesco Berto. I am grateful to them for helping me to develop my thinking on these issues. I particularly want to thank Ferenc Csaba for his prepared response to my lecture at the Budapest meeting and to thank all those with whom I have discussed these issues and those who have raised issues in response to my various presentations or to drafts of this paper. These include Daniel Andler, George Boolos, Christopher Broadbent, Harvey Brown, John Burgess, Oswaldo Chateaubriand, Karine Chemla, Peter Clark, Martin Davis, Philippe de Rouilhan, Jacques Dubucs, Michael Dummett, Roy Dykoff, Solomon Feferman, Michael Gabbay, Alexander George, Volker Halbach, Russell Howell, Jeffrey Ketland, Jonathan Kirby, Elaine Landry, Hannes Leitgeb, Øystein Linnebo, John Macfarlane, Penelope Maddy, John Mayberry, Koji Nakatogawa, Charles Parsons, Alexander Paseau, Richard Pettigrew, Graham Priest, Ian Rumfitt, Simon Saunders, Dana Scott, Stewart Shapiro, Wilfried Sieg, Peter Smith, Robert Thomas, Gabriel Uzquiano, Kenneth Wächter, Bruno Whittle, Robbie Williams, Timothy Williamson, Crispin Wright. I apologize to those I have not remembered and to those I have

remembered but whose points I have not properly taken into account (either because I am being stubborn or obtuse or simply have not found the time to rewrite my text on that point). I am tremendously grateful to Georg Kreisel for discussing these issues with me (which does not mean that he agrees with anything I have written here, including when I take myself to be expounding and endorsing his views), and to Alex Wilkie for his always clear answers and endless patience in response to my many questions about the results in set theory on which the philosophical ideas in this paper depend, and also to Philip Welch, Hugh Woodin, and Boris Zilber for patient explanations. The degree to which I am critical of the views of Stewart Shapiro is a measure of the extent of my indebtedness to his writings on structuralism and to his generosity in discussion. I am grateful to Oxford University for sabbatical leave during which I developed these ideas and to the Arts and Humanities Research Council for a Research Leave Award that gave me further time in which to work on this project. I am deeply grateful to my wife Kassandra for her love and encouragement while I have worked on this topic. I dedicate this paper to our son Jonathan, who was born during the time when I began to think about the implications of quasi-categoricity of set theory for philosophy of mathematics.

References

- Aczel, P. 1988. *Non-Well-Founded Sets*. CSLI Lecture Notes no. 14. Stanford, Calif.: Center for the Study of Language and Information.
- Bell, J.L. 1969. *Model Theoretic and Axiomatization Results for Certain Restricted Second Order Languages*. D.Phil. thesis, University of Oxford.
- Benacerraf, P. 1965. "What Numbers Could Not Be." *Philosophical Review* 74: 47–73.
- . 1973. "Mathematical Truth." *Journal of Philosophy* 70: 661–680.
- . 1996. "Recantation." *Philosophia Mathematica* 4: 184–189.
- . 1999. "What Mathematical Truth Could Not Be—II, or Through a Glass Darkly." In *Sets and Proofs: Invited Papers from Logic Colloquium '97—European Meeting of the Association of Symbolic Logic, Leeds, July 1997*, edited by S.B. Cooper and J.K. Truss, 27–51. London Mathematical Society Lecture Notes Series 258. Cambridge: Cambridge University Press.
- Bernays, P. 1930. "Die Philosophie der Mathematik und die Hilbertsche Beweistheorie." *Blätter für deutsche Philosophie* 4: 326–367. Reprinted with postscript in *Abhandlungen zur Philosophie der Mathematik*. Darmstadt:

- Wissenschaftliche Buchgesellschaft, 1976. English translation by Paolo Mancosu as "The Philosophy of Mathematics and Hilbert's Proof Theory." Text no. 9. Bernays Project, Carnegie-Mellon University.
- . 1967. "What Do Some Recent Results in Set Theory Suggest?" In Lakatos 1965, vol. 1, 109–117.
- Boolos, G. 1971. "The Iterative Conception of Set." *Journal of Philosophy* 68: 215–231.
- . 1975. "On Second-Order Logic." *Journal of Philosophy* 72: 509–527.
- . 1998. "Must We Believe in Set Theory?" In *Logic, Logic, and Logic*, 120–132. Cambridge, Mass.: Harvard University Press.
- Bourbaki, N. 1948. "L'architecture des mathématiques." In *Les Grands Courants de la Pensée Mathématique*, edited by F. Le Lionnais. Cahiers du Sud, 35–47. English translation by A. Dresden as "The Architecture of Mathematics." *American Mathematical Monthly* 57 (1950): 221–232.
- Burgess, J. and R. Gideon. 1997. *A Subject with no Object: Strategies for Nominalistic Interpretation of Mathematics*. New York: Oxford University Press.
- Cantor, G. 1878. "Ein Beitrag zur Mannigfaltigkeitslehre." *Journal für die reine und angewandte Mathematik* 84: 242–258.
- . 1899. Letter to R. Dedekind, 28 July 1899. In Cantor 1932, 443–447. English translation by S. Bauer-Mengelberg and J. van Heijenoort in van Heijenoort 1967, 113–117.
- . 1932. *Gesammelte Abhandlungen matematischen und philosophischen Inhalts*. Edited by E. Zermelo. Berlin: Springer.
- Cohen, P.J. 1963. "The Independence of the Continuum Hypothesis I." *Proceedings of the National Academy of Sciences of the U.S.A.* 50: 1143–1148.
- . 1964. "The Independence of the Continuum Hypothesis II." *Proceedings of the National Academy of Sciences of the U.S.A.* 51: 105–110.
- Cohen, P.J. and R. Hersh. 1967. "Non-Cantorian Set Theory." *Scientific American* 217: 104–116.
- Corry, L. 1992. "Nicolas Bourbaki and the Concept of Mathematical Structure." *Synthese* 92: 315–348.
- . 1998. "The Origins of Eternal Truth in Modern Mathematics: Hilbert to Bourbaki and Beyond." *Science in Context* 12: 137–183.
- Darmon, H. 1999. "A Proof of the Full Shimura-Taniyama-Weil Conjecture Is Announced." *Notices of the American Mathematical Society* 46: 1397–1401.
- Dedekind, R. 1872. *Stetigkeit und irrationale Zahlen*. Braunschweig: Vieweg. English translation by W.W. Beman as "Continuity and Irrational Numbers," in Dedekind 1901, 765–779. Reprinted with corrections in Ewald 1996, 765–779.
- . 1888. *Was sind und was sollen die Zahlen?* Braunschweig: Vieweg. 2nd ed. 1893. English translation by W.W. Beman as "The Nature and Meaning of Numbers," in Dedekind 1901, 29–115. Extensively revised by W.B. Ewald, in Ewald 1996, 787–833.
- . 1901. *Essays on the Theory of Numbers*. Chicago: Open Court.
- Drake, F. 1974. *Set Theory: An Introduction to Large Cardinals*. Amsterdam: North-Holland.

- Ewald, W.B., ed. 1996. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*. New York: Oxford University Press.
- Enderton, H.B. 1977. *Elements of Set Theory*. New York: Academic Press.
- Feferman, S. 1984. "Kurt Gödel: Conviction and Caution." *Philosophia Naturalis* 21: 546–562. Reprinted with additions and corrections in *In the Light of Logic*, 150–164. New York: Oxford University Press, 1998.
- . 1999. "Does Mathematics Need New Axioms?" *American Mathematical Monthly* 106: 99–116.
- . 2000. "Why the Programs for New Axioms Need To Be Questioned." *Bulletin of Symbolic Logic* 6: 401–413.
- Field, H. 2000. "Indeterminacy, Degree of Belief, and Excluded Middle." *Noûs* 34: 1–30.
- Frege, G. 1884. *Die Grundlagen der Arithmetik, eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Breslau: Verlag von Wilhelm Koebner. English translation by J.L. Austin as *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*. Oxford: Basil Blackwell, 1950. 2nd ed. 1953. Revised ed. 1980.
- . 1976. *Wissenschaftlicher Briefwechsel*. Edited by G. Gabriel et al. Hamburg: Felix Meiner Verlag. Abridged by B. McGuinness and translated into English by H. Kaal as *Philosophical and Mathematical Correspondence*. Oxford: Basic Blackwell, 1980.
- Friedman, H. 1971. "Higher Set Theory and Mathematical Practice." *Annals of Mathematical Logic* 2: 326–357.
- Gödel, K. 1931. "Über formal unentscheidbare Sätze der *Principia mathematica* und verwandter Systeme I." *Monatshefte für Mathematik und Physik* 38: 173–198. English translation by J. van Heijenoort as "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems I," in *Publications 1929–1936*, vol. 1 of *Collected Works*, edited by S. Feferman et al., 144–195. New York: Oxford University Press, 1986.
- . 1947. "What is Cantor's Continuum Problem?" *American Mathematical Monthly* 54: 515–525. Errata 55: 151. Revised and expanded version in *Philosophy of Mathematics: Selected Readings*, edited by P. Benacerraf and H. Putnam, 258–273. Oxford: Basil Blackwell, 1964. Reprinted in Gödel 1990, 176–187, 254–270.
- . 1959. Letter to Paul Arthur Schilpp, 3 February 1959. In Gödel 2003, 244.
- . 1965. "Remarks Before the Princeton Bicentennial Conference on Problems in Mathematics." In *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvability Problems and Computable Functions*, edited by M. Davis, 84–88. New York: Raven Press, Hewlett. Reprinted in Gödel 1990, 150–153.
- . 1970. "Some Considerations Leading to the Probable Conclusion That the True Power of the Continuum Is $\aleph_2$ " (1970a). "A Proof of Cantor's Continuum Hypothesis from a Highly Plausible Axiom about Orders of Growth" (1970b). "Unsent letter to Alfred Tarski" (1970c). Introductory note to 1970a, 1970b and 1970c by R.M. Solovay. In Gödel 1995, 405–425.

- . 1990. *Publications 1938–1974*. Vol. 2 of *Collected Works*, edited by S. Feferman et al. New York: Oxford University Press.
- . 1995. *Unpublished Essays and Lectures*. Vol. 3 of *Collected Works*, edited by S. Feferman et al. New York: Oxford University Press.
- . 2003. *Correspondence H–Z*. Vol. 5 of *Collected Works*, edited by S. Feferman et al. New York: Oxford University Press.
- Gödel K. and E. Zermelo. 1931. Correspondence. Printed with English translation by J.W. Dawson in Gödel 2003, 420–431.
- Goldfarb, W. 1995. Introductory note to drafts of “Is Mathematics Syntax of Language?” by K. Gödel. In Gödel 1995, 324–334.
- Goodstein, R.L. 1944. “On the Restricted Ordinal Theorem.” *Journal of Symbolic Logic* 9: 33–41.
- Gowers, T. 2006. “Does Mathematics Need a Philosophy?” In *18 Unconventional Essays on the Nature of Mathematics*, edited by R. Hersh, 182–200. New York: Springer.
- Hazen, A.P. 1985. Review of *Frege’s Conception of Numbers as Objects*, by C. Wright. *Australasian Journal of Philosophy* 63: 253–254.
- Hellman, G. 1989. *Mathematics without Numbers: Towards a Modal-Structural Interpretation*. New York: Oxford University Press.
- Hilbert, D. 1899. *Grundlagen der Geometrie*. Leipzig: Teubner. English translation by E.J. Townsend as *Foundations of Geometry*. Chicago: Open Court, 1902. 10th ed., revised and enlarged by Paul Bernays. Stuttgart: Teubner, 1968. English translation by Leo Unger of the 10th ed. La Salle, Ill.: Open Court, 1971.
- Isaacson, D. 1987. “Arithmetical Truth and Hidden Higher-order Concepts.” In *Logic Colloquium ’85: Proceedings of the Colloquium held in Orsay, France July 1985*, edited by The Paris Logic Group, 147–169. Amsterdam: North-Holland. Reprinted with small revisions in *The Philosophy of Mathematics*, edited by W.D. Hart, 201–224. New York: Oxford University Press, 1996.
- . 1992. “Some Considerations on Arithmetical Truth and the ω -rule.” In *Proof, Logic and Formalization*, edited by M. Detlefsen, 94–138. London: Routledge.
- . 1994. “Mathematical Intuition and Objectivity.” In *Mathematics and Mind*, edited by A. George, 118–140. New York: Oxford University Press.
- . 2004. “Quine and Logical Positivism.” In *The Cambridge Companion to Quine*, edited by R. Gibson, 214–269. Cambridge: Cambridge University Press.
- Jech, T. 2003. *Set Theory: The Third Millennium Edition, revised and expanded*. Berlin: Springer.
- Kalmár, L. 1965. “On the Role of Second Order Theories.” In Lakatos 1965, vol. 1, 104–105.
- Kanamori, A. 1997. *The Higher Infinite: Large Cardinals in Set Theory from Their Beginnings*. Berlin: Springer-Verlag, 1997.

- Katz, J.J. 1996. "Skepticism about Numbers and Indeterminacy Arguments." In *Benacerraf and his Critics*, edited by A. Morton and S.P. Stich, 119–139. Oxford: Blackwell Publishers.
- . 1998. *Realistic Rationalism*. Cambridge, Mass.: MIT Press.
- Keränen, J. 2001. "The Identity Problem for Realist Structuralism." *Philosophia Mathematica* 9: 308–330.
- Kirby, L. and J. Paris. 1982. "Accessible Independence Results for Peano Arithmetic." *Bulletin of the London Mathematical Society* 14: 285–293.
- Kluge, E.-H.W., ed., trans. 1971. *Gottlob Frege on the Foundations of Geometry and Formal Theories of Arithmetic*. New Haven: Yale University Press.
- Kreisel, G. 1950. "Note on Arithmetic Models for Consistent Formulae of the Predicate Calculus." *Fundamenta Mathematicae* 37: 241–267.
- . 1967. "Informal Rigour and Completeness Proofs." In Lakatos 1965, vol. 1, 138–171.
- . 1968. "A Survey of Proof Theory." *Journal of Symbolic Logic* 33: 321–388.
- . 1977. Review of Weston 1976. *Zentralblatt für Mathematik* 341: 31.
- Lakatos, I., ed. 1965. *Problems in the Philosophy of Mathematics: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*. 3 vols. Amsterdam: North-Holland Publishing Co.
- Lang, S. 1985. *The Beauty of Doing Mathematics: Three Public Dialogues*. New York: Springer-Verlag.
- Levy, A. 1979. *Basic Set Theory*. Berlin and New York: Springer-Verlag.
- Linnebo, Ø. 2008. "Structuralism and the Notion of Dependence." *Philosophical Quarterly* 58: 59–79.
- Lopez-Escobar, E.G.K. 1967. "A Complete, Infinitary Axiomatization of Weak Second-Order Logic." *Fundamenta Mathematicae* 61: 93–103.
- MacBride, F. 2005. "Structuralism Reconsidered." In *The Oxford Handbook of Philosophy of Mathematics and Logic*, edited by S. Shapiro, 563–589. New York: Oxford University Press.
- Manevitz, L.M. and J. Stavi. 1980. "Operators and Alternating Sentences in Arithmetic." *Journal of Symbolic Logic* 45: 144–154.
- Martin, D.A. 1975. "Borel Determinacy." *Annals of Mathematics* 102: 363–371.
- . 1976. "Hilbert's First Problem: The Continuum Problem." In *Mathematical Developments Arising From Hilbert Problems: Proceedings of Symposia in Pure Mathematics* 28, edited by F. Browder, 81–92. Providence, Rhode Island: American Mathematical Society.
- Martin, D.A. and J.R. Steel. 1988. "Projective Determinacy." *Proceedings of the National Academy of Sciences of the USA* 85: 6582–6586.
- Martin, R.M. 1963. *Intension and Decision: a Philosophical Study*. Englewood Cliffs, N.J.: Prentice-Hall.
- McGee, V. 1997. "How We Learn Mathematical Language." *Philosophical Review* 106: 35–68.
- Monk, R. 1990. *Ludwig Wittgenstein: The Duty of Genius*. London: Jonathan Cape.

- Moore, G. 1980. "Beyond First-Order Logic: The Historical Interplay between Mathematical Logic and Axiomatic Set Theory." *History and Philosophy of Logic* 1: 95–137.
- . 1990. Introductory note to "What is Cantor's Continuum Problem." In Gödel 1990, 154–175.
- Mostowski, A. 1967. "Recent Results in Set Theory." In Lakatos 1965, vol. 1., 82–96.
- Parsons, C. 1990. "The Structuralist View of Mathematical Objects." *Synthese* 84: 303–346.
- Prestel, A. 1990. "Nonstandard Analysis." In *Numbers*, edited by J.H. Ewing, 305–327. New York and London: Springer-Verlag.
- Poincaré, H. 1902. "La grandeur mathématiques et l'expérience." *La Science et l'Hypothèse*. Bibliothèque de Philosophie Scientifique. Paris: Ernest Flammarion. English trans. by W.J. Greenstreet as "Mathematical Magnitude and Experiment." In *Science and Hypothesis*, 17–34. London: Walter Scott Publishing Co, 1905.
- Putnam, H. 1967. "The Thesis that Mathematics is Logic." In *Bertrand Russell: Philosopher of the Century*, edited by R. Shoenman, 273–303. London: George Allen and Unwin. Reprinted in *Mathematics, Matter and Method*, vol. 1. of *Philosophical Papers*, 12–42. Cambridge: Cambridge University Press, 1975.
- Quine, W.V. 1970. *Philosophy of Logic*. Englewood Cliffs, N.J.: Prentice-Hall. 2nd edition. Cambridge, Mass: Harvard University Press, 1986.
- Resnik, M. 1975. "Mathematical Knowledge and Pattern Cognition." *Canadian Journal of Philosophy* 5: 25–39.
- . 1988. "Mathematics from the Structural Point of View." *Revue Internationale de Philosophie* 42: 400–424.
- Russell, B. 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.
- Shapiro, S. 1985. "Second-Order Languages and Mathematical Practice." *Journal of Symbolic Logic* 50: 714–742.
- . 1997. *Philosophy of Mathematics: Structure and Ontology*. New York: Oxford University Press.
- . 2008. "Identity, Indiscernibility, and *ante rem* Structuralism: The Tale of *i* and *-i*." *Philosophia Mathematica* 16(3): 285–309.
- Shepherdson, J.C. 1951–1953. "Inner Models for Set Theory." Parts 1, 2, 3. *Journal of Symbolic Logic* 16: 161–190; 17: 225–237; 18: 145–167.
- Smoryński, C. 1977. "The Incompleteness Theorems." In *Handbook of Mathematical Logic*, edited by J. Barwise, 821–865. Amsterdam: North-Holland Publishing Co.
- van Heijenoort, J., ed. 1967. *From Frege to Gödel: A Source Book in Mathematical Logic, 1897–1931*. Cambridge, Mass.: Harvard University Press.
- Väänänen, J. 2001. "Second-Order Logic and Foundations of Mathematics." *Bulletin of Symbolic Logic* 7: 504–520.
- Weston, T. 1976. "Kreisel, the Continuum Hypothesis and Second Order Set Theory." *Journal of Philosophical Logic* 5: 281–298.

- Woodin, W.H. 2001. "The Continuum Hypothesis." *Notices of the American Mathematical Society* 48: 567–576, 681–690.
- . 2005. "Set Theory after Russell: The Journey back to Eden." In *One Hundred Years of Russell's Paradox: Mathematics, Logic, Philosophy*, edited by G. Link, 29–47. De Gruyter Series in Logic and its Applications, vol. 6. Berlin: de Gruyter.
- Zermelo, E. 1908. "Untersuchungen über die Grundlagen der Mengenlehre I." *Mathematische Annalen* 59: 261–281. English translation by S. Bauer-Mengelberg in van Heijenoort 1967, 199–215.
- . 1930. "Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre." *Fundamenta Mathematicae* 14: 29–47. English translation by M. Hallett as "On Boundary Numbers and Domains of Sets: New Investigations in the Foundations of Set Theory," in Ewald 1996, 1209–1233.
- . 1932. Editor's Remark on Cantor's Review of Frege's *Die Grundlagen der Arithmetik*. In Cantor 1932, 441–442.

NENAD MIŠČEVIĆ

Conceptualism and Knowledge of Logic

A Budget of Problems

1 Introduction

Ordinary cognizers reason, at least sometimes, in accordance with logical rules. They find some instances of logical principles compelling and obvious, provided the latter are sufficiently simple and undemanding. They are “sensitive to logical form,” as Russell would put it.¹ Explicit elementary knowledge of logic probably derives from these simple sensitivity and abilities. How is the elementary knowledge of logic, both implicit and explicit, justified or warranted? What entitles the cognizers use of logical principles in inference? And, assuming a broadly realist stance about logic, what justifies the assumption that these principles are reliable and objectively valid? How do we access “logical reality” (so to speak), if it is objective and mind-independent? The last question suggests a version of Benacerraf’s dilemma for logic, with obvious ties to its original area, philosophy of mathematics. The tie is suggested, among other things, by the prominence and promise of logicism: if our knowledge of mathematics can be epistemologically grounded in our knowledge of logic, in insight into the latter promises also to solve the mysteries of the former. Another tie might be the plausibility of a Platonic ontology of logical entities, propositions

I wish to thank the participants of the Oxford–Budapest conference, the participants of the Russell conference at CEU (and the organizer, Zsófia Zvolenszky), and those of the Rijeka philosophy of logic conference for discussion and help, and most of all the editors of the volume, Zs. Novák and A. Simonyi, for their invitation, encouragement and long and patient struggle with all the mistakes I made in the draft version of the paper.

¹ See Russell 1927, 67–68.

and their logical relations.² Since logic is a paradigmatic example of a (candidate) *a priori* domain, this issue lies at the intersection of at least two lines of inquiry: one concerning the epistemology of logic itself, the other the epistemology of (candidate) *a priori* beliefs and practices of reasoning.

A dominant line in recent debate is a variety of the apriorist or rationalist line, relying upon cognizers mastery of logical concepts: the mastery gives the cognizer access to “logical reality” and thus ultimately helps to solve Benacerraf’s dilemma. I shall call it “conceptualism,” taking as its main proponents C. Peacocke and P. Boghossian. I will also appeal to the work of B. Hale and C. Wright, since it comes quite close to the mainstream conceptualist position. Here I am interested in specifically epistemological issues, so by “conceptualism” I will, in the sequel, mean above all, an epistemological position. It rejects any reliance on *a posteriori* sources of justification or warrant for logical knowledge, mostly rejects or downplays the need of any kind of broadly causal (or causal-like) explanation of our logical capacities, both of our having them and of their reliability. Instead, the authors listed proposed a combination of three sources of justification. The first one is the alleged (meaning-or-concept) constitutive work that logical principles perform in fixing the sense and semantic value of logical expressions (above all constants). This is a logic-immanent, concept-focused, kind of justification: I shall call it justification-through-constitution. The two others come from cognitive-epistemic considerations, somewhat external to logic itself. The main candidates are first, obviousness-cum-compellingness, and second, indispensability, that allegedly makes logical principles into a kind of universal hinge-propositions (“cornerstones,” as Wright calls them; see below). The three sources are taken to be of purely *a priori* nature, and standing in no need of broadly empirical explanation of an ordinary causal kind, for instance psychological or evolutionary.

In this paper I would like to argue against these last two assumptions, and in favor of deploying some *a posteriori* considerations together with a modicum of explanationism in reflectively justifying logical knowledge. I agree that logic is justified immediately and in this weak sense *a priori*, but only in a relatively *prima facie* and unreflective

² Thanks go to the two editors for reminding me of this second tie.

way. A deeper skeptical probing, and a more demanding and reflective drive for justification and understanding brings in *non-a priori* components: *there is nothing wrong in reflectively justifying our logical beliefs and habits by methods of wide reflective equilibrium, taking into account both their stunning empirical success, and the available empirical explanation(s) of cognizers having them and of their reliability. The three conceptualist considerations should be supplemented (and, some might argue, even replaced) by wide reflective equilibrium at the reflective level: first, the meaning-constitutive account moves in a too small circle of a very narrow reflective equilibrium, so there is a need to expand the circle, which would make the equilibrium wide. Second, the logic-external consideration of obviousness-cum-compellingness cries for explanation. And finally, the logic-external consideration of indispensability leads directly to empirical considerations having to do with success, actual or potential, of the cognitive enterprise(s) for which logic is so badly needed. And success is an indicator of reliability.*³ In short, staying within the narrow conceptualist circle amounts to placing of a “veil of conception” between us and the logic-in-the world, blocking the understanding of what makes logic objectively valid. Thus, the very dialectics of the conceptualist program points to a less apriorist view of logical knowledge. This is the view to be defended indirectly in the paper, by offering an overview of conceptualist program and a budget of problems for it. Epistemological conceptualism about logic has not been much discussed as a unitary epistemological position, although particular claims of particular authors have been rather thoroughly examined.⁴ The present paper summarizes what appear to be the main difficulties for it. I hope that brevity might be excused by a gain in the

³ The editors Zs. Novák and A. Simonyi have made me aware that Quine is often read as a deflationist concerning both truth and logic; they usefully contrast him with tough realists like Alvin Goldman. Their proposal is reliability/realism/metaphysical naturalism/causal access (Goldman) versus empirical success/indispensability/deflationism/methodological naturalism/pragmatist holism (Quine), in which context they mention incompatible naturalist traditions. My own reading of Quine himself is much more realist (Mišćević 2000), but I will not enter Quine scholarship here. I shall therefore speak mostly of Quine-Putnam line, having in mind early, realist Putnam and his development of realistic indispensability arguments for logic and mathematics.

⁴ Williamson's very thorough paper on conceptual truth (Williamson 2006), to which I refer later, might be a beginning of the change in this respect, but he concentrates more upon semantic aspect, and treats epistemological matters as consequences.

bigger picture which can make the systematic nature both of the program and of its difficulties clear.

Before passing to the articulation of the paper, let me briefly sketch my preferred epistemological framework. Traditionally, the discussion in the epistemology of logic has focused upon reasons in principle accessible to the cognizer herself. This internalist bent has survived all the attractions of externalism; since conceptualists to be discussed all share it, I will join them in focusing upon such reasons. However I think, following in this B. Russell, that we have to account both for the external reliability of our “sensitivity to logical form,” and for our internally accessible reasons (see Russell 1940, 12). If a cognitive capacity is reliable thus having an external justification, and the cognizer also has available to herself a good reason to trust, thereby having internal justification, then she is justified in an absolute sense. So, on the externalist side, reliability is the ultimate virtue of our logical ability. However, I shall limit myself here only to raising the standard question of how the reliability is to be explained, without going into any further detail. In contrast, I concentrate here upon justification (warrant, entitlement) in principle accessible to the cognizer, also discussed by my conceptualist opponents. I am attracted to a rich picture, also inspired by Russell, that combines two elements: immediate, *prima facie* justification by obviousness-compellingness, and a more coherentist one, more suited for reflective justification.⁵ The most interesting issue for logical beliefs and inferential propensities is the issue of the kind of coherence, or reflective equilibrium that is relevant, i.e. the choice between narrow and wide versions of equilibrium.⁶ Besides arguing for the wide one, I will be stressing the point that providing an explanatory account is an integral part of full reflective justification. Such an account can explain the having of logical-mathematical

⁵ Russell presents this two-component view of epistemology in his latter works, starting from *The Analysis of Mind* (1921), through *An Inquiry into Meaning and Truth* (1940), up to *Human Knowledge: Its Scope and Limits* (1948), encompassing both internalist and externalist elements. There are two different inquiries, both important, and each having a right to the name “theory of knowledge.” In any given discussion, he says, it is easy to fall into confusions through failure to determine to which of the two inquiries the discussion is intended to belong (Russell 1940, 12).

⁶ E. Sosa, who also advocates a rich picture, beyond the “false dichotomies” of internalism versus externalism and foundationalism versus coherentism (Sosa 2004), suggests that a narrow reflective equilibrium is sufficient (Sosa 2003).

intuitions, their reliability, as well as mistakes that sometimes accompany them. It can even, as Casullo (2003) has argued tell us which, if any, of our beliefs are justified *a priori*. We shall see in the sequel that the main conceptualist apriorist we discuss, routinely takes explanation to play an important role in our epistemological reflection when it comes to *a posteriori* beliefs; so why not for all beliefs. The awareness of the link between explanation and justification has traditionally fuelled the interest in the epistemology of mathematics (and later of logic) from Plato through Descartes, Kant and Russell to contemporaries like Benacerraf. Note that the two last decades of the debate in mathematical epistemology have been fueled by Benacerraf's dilemma, crucially featuring the notion of causal explanation. The guiding idea is that as long as we don't have an explanatory story, our reflective justification will be defective.⁷ Of course, even if we do not have it, we can still be weakly justified, say, in implicitly using logic in our reasoning. However, the full internal reflective justification will be missing. And the external reliability might just be a matter of good luck for all we reflectively know (a bit of change and our procedures would not track truths any more). Note the contrast with the case of perception where we seem to have it: part of our confidence in it is that we sort of understand at the level of common sense how material things can act upon us, and that, further, we have some scientific assurance that this commonsense understanding is on the right track. For these reasons I think that explanation is a component of wider reflective equilibrium,

⁷ The relevant states are belief states of some kind, so they must be caused some way or another (if you are analytic functionalist, this is for you a conceptual truth). Then, the way they are caused they are either suitably connected to their content (e.g. mathematical intuitions might result from cognitive makeup that has been evolutionary reared on instantiations of mathematical properties in the actual world), or *they are not connected to the content in any way*. If the latter, *then our logical and mathematical capacity and related belief-states seem to be totally unconnected to states of affairs their contents are about, but causally connected to completely irrelevant antecedents*. That horn does not preserve purity of intuitions because there are enough content-irrelevant causal antecedents to "spoil" it; there is a plethora of competing content-irrelevant causes. Neither does it offer nor preserve certainty: why would one trust a capacity whose deliverances are explainable by completely irrelevant causal routes? Indeed, it makes things much worse than any standard naturalistic explanation. The first, i.e. compatibilist horn, featuring indirect connection of intuition capacity and state with some instantiations of properties presented by their contents is much more soothing for our fears and worries, but it is fatal to the incompatibilists.

and that this fact could undermine serious apriority of central armchair pieces of knowledge: causal explanation is *a posteriori*, therefore, justification might be partly *a posteriori*.

To reiterate, this paper offers an overview of the conceptualist approach, discusses three apriorist strategies of justifying logical beliefs and inferential practices, prominent in the recent conceptualist wave, first, the one from constitutiveness, second from obviousness-cum-compellingness and third from indispensability. Here is the preview: the balance of the introduction offers a short summary of the background to the debate, concerning the relation of logic and natural reasoning, the assumption of realism as against anti-realism, and stresses the task of justifying our belief in objective validity of (core) logic. Two prominent ways of defending the objectivity of logic are mentioned, the Quine-Putnamian *a posteriori* way from indispensability and “versatility” of logic (where I take Putnam’s realist continuation of Quine as the guiding thread), and the more traditional insight-based, direct referential view exemplified by BonJour. Section 2, “Conceptualism: constitution, compellingness and obviousness” discusses the two conceptualist strategies mentioned in its title, taking Peacocke’s proposal as its focus. It first places recent conceptualism in relation to these two alternatives, and then passes to the two strategies of defending logical beliefs: the one from their constitutive role, and the other from their obviousness and compelling character. Since Peacocke is offering an impressive combined defense, its two prongs are here presented together, and discussed in the same section. The discussion focuses upon the narrow circle of constitution, the issue of explaining the compelling character of logical knowledge and the issue of explanation in general. Section 3 “Indispensability” turns to the third strategy, defended by C. Wright (2004b). It contrasts his apriorist interpretation of indispensability and the classical Quinean-Putnamian one, and argues that there is no ground to believe that indispensability guarantees apriority. Section 4 recapitulates the budget of problems for conceptualism. In short, understanding logical constants might not be constitutive for having the concept that is its sense, there are problems of circularity involved in justifying and there is the problem of determination, a variant of Benacerraf’s dilemma. The classical problem for obviousness-cum-compellingness is the dilemma between a psychological and a normative-epistemological

interpretation and there are skeptical problems targeting the same justifier. The appeal to indispensability subverts the *a priori* nature of justification or warrant involved since it brings into the play the empirical considerations having to do with success, actual or potential, of the cognitive enterprises for which logic is indispensable. Given all these problems, a more explanation-friendly alternative might be preferable, and it is probable that such an alternative will give more weight to *a posteriori* considerations in justification.

Finally, an apology. Writing about conceptualism in general, without singling out particular authors brings the risks of being accused of attacking a straw man.⁸ Singling out an author or authors requires going into specific details. I have tried to find a middle way, treating Peacocke and Wright as representatives of more general strategies; I am aware of running a risk of combining and stressing only bad sides of various methods proposed by various authors: appearing superficial and unjust to authors selected, and still too specific and dismissive of other possibilities explored by other conceptualists. But this is the risk that is hard to avoid given the sheer complexity and richness of the extant literature.

Let me then very briefly sketch the background, namely the relevant understanding of “logical knowledge,” and the range of options available for accounting for it. Following the usage in most of contemporary debate, I am using “logical knowledge” without qualification in two senses. First, as a wide, umbrella term covering all sorts of cognitive mastery having to do with valid deductive inference, from the initial implicit, unsophisticated and very fallible (often simply erroneous) elementary one to the highly sophisticated explicit knowledge of a professional logician. Second, in the sense more focused upon the former, spontaneous and elementary variety, which I take to be cognitively fundamental. I shall rely upon context to specify the intended sense. Further, following the tradition, I shall often slightly abuse the term “knowledge” to cover correct beliefs as well as inferential steps, and indeed in the wide sense encompassing both standing and occurrent varieties of each: standing beliefs, occurrent beliefs, standing inferential dispositions and performed inferences.

⁸ For example, Claire Jenkins (on her blog *Long Words Bother Me* [2 July, 2006]) is accusing T. Williamson (2006) of such a fault.

Let us start with the last group. The best illustration of it is offered by those cases in which the cognizer spontaneously and implicitly recognizes an instance of logical consequence.⁹ So, suppose Dorothy has been told by her daughter that their dog has a broken leg and a deep laceration. She quickly passes from this alarming information to the decision to quickly get some bandages, one intermediate, implicit step taking her from the thought “the dog has a broken leg and a deep laceration” to the thought “the dog has a deep laceration.” Her implicit handling of conjunction elimination can be brought to light by explicitly asking her a question, e.g. whether it is possible that the dog has a broken leg and a deep laceration, without having a broken leg. The “of course, not” answer would confirm the impression that she does have a mastery of the rule governing conjunction. Call this knowledge instance-knowledge. It seems that knowledge manifested in such spontaneous inferences is knowledge how. But, when a naïve thinker begins to reflect, the inferential step seems obvious, but not in the way the logic manual presents it. In our example, it is the impossibility of the combined situation itself, a-dog-having-a-broken-leg-and-a-deep-laceration-without-having-a-broken-leg that appears compelling to Dorothy. She is thinking of (what we, theoreticians would call) *semantic values* of the (Fregean) propositions in question, not of the propositions-senses themselves. And the impossibility of the combined concrete situation is obvious and is luminously presented to her. Thus she finds the situation in which it is the case that p -and- q but not- p (for some instances of “ p ” and “ q ”) inconceivable. It is not something about the sentence or thought itself that is inconceivable, let alone about some formula

⁹ Of course, for elementary moves, like conjunction elimination or Modus Ponens, the inference is not explicit. Russell even claims that the inference is rarely performed at all, as Russell notes (1927, 82):

This form of inference does actually occur, though very rarely. The only instance I have ever heard of was supplied by Dr. F.C.S. Schiller. He once produced a comic number of the philosophical periodical *Mind*, and sent copies to various philosophers, among others to a certain German, who was much puzzled by the advertisements. But at last he argued: “Everything in this book is a joke, therefore the advertisements are jokes.” I have never come across any other case of new knowledge obtained by means of a syllogism. It must be admitted that, for a method which dominated logic for two thousand years, this contribution to the world’s stock of information cannot be considered very weighty.

employing schematic letters. Russell rightly thought that such spontaneous inference counts as recognizing the logical form of conjunction and talked of “acquaintance with” and “sensitivity to” logical form (Russell 1927, 87–88).¹⁰ Given the complications of ‘acquaintance,’ we shall use more the latter expression. In short, logical knowledge begins with knowledge of instances of logical relations, and instances of validity of logical principles. It is manifested in simple inference.

The description just given accords with the most natural reading of conceptualists themselves, in particular Peacocke (see References); they are not very explicit about concrete examples, but the reading certainly does not beg any questions about their views, and is fair if not indeed charitable; we shall return to it when discussing Peacocke. The description seems natural to many, but it does take a substantial stance on the issue of the relation between logic and ordinary reasoning and psychological and epistemic status of logic. It claims that some deductive logic is a part of the natural reasoning competence, and is

¹⁰ He explains it as an instance of the general “power of reacting to form” and adds that it is the capacity that “chiefly characterizes ‘intellect’” (Ibid.). But he notes that some of the higher animals also possess it, though to nothing like the same extent as men do; and all animals except a few of the most intelligent species appear to be nearly devoid of it. Logical form is just a special case of form. Russell also saw a continuity between acquaintance with logical form and with more “ordinary” universals, like red. He takes a similar approach, now from the externalist standpoint.

When a child is being taught to read, he learns to recognize a given letter, say H, whether it is large or small, black or white or red. However it may vary in these respects his reaction is the same: he says “H.” That is to say, the essential feature in the stimulus is its form. When my boy, at the age of just under three, was about to eat a three-cornered piece of bread and butter, I told him it was a triangle. (His slices were generally rectangular.) Next day, unprompted, he pointed to triangular bits in the pavement of the Albert Memorial, and called them “triangles.” Thus the form of the bread and butter, as opposed to its edibility, its softness, its color, etc., was what had impressed him. This sort of thing constitutes the most elementary kind of reaction to form. (88)

The rest is easy, Russell suggests:

To “understand” even the simplest formula in algebra, say $(x+y)^2 = x^2 + 2xy + y^2$, is to be able to react to two sets of symbols in virtue of the form which they express, and to perceive that the form is the same in both cases. This is a very elaborate business, and it is no wonder that boys and girls find algebra a bugbear. But there is no novelty in principle after the first elementary perceptions of form. (89)

therefore incompatible with the radical view according to which deductive logic is an invention unrelated to the ordinary functioning of our cognitive system, which might be described as cognitive anti-logicism. If the view were correct we would need completely separate justificatory accounts for ordinary deduction-like inference and deductive logic. Here I shall just assume the incorrectness of the view. The description given is compatible with two kinds of pro-logical views: the more demanding one according to which logic (classical logic) is the core of our natural cognitive system: this is full epistemic logicism familiar from the writings of Lance Ripps (1994). The second is that logic is a part of our system (but not necessarily its core); call it minimal logicism. Both allow for common justification for logic and (some parts or features of) ordinary rationality. To reiterate, the continuity between the common reasoning competence and professional logic is a shared assumption and a common ground with conceptualists.

One might object at the present juncture, as the editors of the present volume did, that my insisting on (the imaginability of) worldly situations as guides to reasoning is excessive. There is no need for “logic in the world.” The short response is the following: on the side of naïve reasoner, the focus upon the imaginability of worldly situations is the only correct one. She does not look into logic book, nor does she consult her memories about the meaning of logical words-constants. She thinks and talks about the world, and our account is faithful to this fact. It is true that logic governs fiction as well as reality, but so does physics, for most fictional works. This does not make physics non-worldly. Equally, the topic-neutrality of logic makes it valid *everywhere*, not nowhere, as both Frege and Quine were famously never tired of stressing. On the side of sophisticated theoretician, various responses are possible, including the anti-realist one. However, the most interesting versions of conceptualism, notably Peacocke’s, are quite objectivist about logic, so objectivism is the common ground with most of the authors criticized, (the only exception is Wright, and I do not discuss his metaphysics of logic, but only his non-logic based arguments for apriority of logical knowledge). I shall also take into account occasional lapses from objectivism, and point to them, in all fairness. The authors examined accept the substantial character of logical principles and rules. For Peacocke these principles belong to a Fregean realm of sense, which then quite objectively determines what happens in the

referential domain. Distinguish three broad ways for a claim to be substantial. First, epistemically, in terms of novelty and fruitfulness of beliefs it expresses. Second, semantically: once their meaning is fixed, *a priori* claims are made true by facts independent of them, and of human mind in general. Finally, metaphysically: some of them concern rich and metaphysically fundamental items, such as, for instance, sets and numbers, or humanly particularly interesting items such as our minds (call this sub-species “anthropologically substantial,” if you want a special name for it). Epistemic substance can probably be accounted for most easily in terms of the limitations of our cognitive grasp, which make logically trivial propositions difficult and exciting for us. However, semantic and metaphysical substance and richness are very difficult to account for. This epistemological task for the realist, namely justifying our belief in objective validity of (core) logic will be the focus of this paper.

Let me remind the reader of the two traditional options, whose weaknesses make up the foil against which conceptualism has asserted itself. The most popular line, in twentieth century, compatible with realism, has been the Quine-Putnamian one according to which logic is *a posteriori*. Quine, who himself vacillates between realism and empiricist anti-realism about logic and science in general, builds his trust in logic and mathematics on a variety of indispensability arguments, relying upon massive empirical success of everyday knowledge and of science in which such beliefs are essentially used. Putnam has given the realistic twist to these arguments. The success, in his view, does vindicate elementary mathematical intuitions, and seems to be relevant for logic as well. The general idea is that every success of science adds a little to the confirmation of its formal, mathematical and logical part.¹¹ On

¹¹ Putnam’s realistic Indispensability Thesis is one instance of the reasoning illustrated here: trust in the truth of mathematics is justified by its indispensability for science. And we crucially need such rationale for the literal truth of mathematical theories. Here is a reconstruction:

(P1) We ought to have ontological commitment to all and only the entities that are indispensable to our best scientific theories.

(P2) Mathematical entities are indispensable to our best scientific theories.

(C) We ought to have ontological commitment to mathematical entities. (Colyvan 2008)

The Indispensability Argument has been famously attacked by Field, but his attack works in our favor, since he rejects both indispensability and truth of mathematics.

the opposite side one finds rationalist-apriorist views, based on direct insight into validity of inference: Laurence Bonjour.

When I carefully and reflectively consider the ... inference ... in question, I am able simply to see or grasp or apprehend that the conclusion of the inference must be true if the premises are true. Such a rational insight, as I have chosen to call it, does not seem to depend on any particular sort of criterion or any further discursive or ratiocinative process, but is instead direct and immediate. (Bonjour 1998, 11)

In finding the idea of rational insight attractive, Bonjour joins a venerable tradition, one that stretches from Plato through Leibniz to Gödel. Most members of the tradition are even more sanguine about insight than Bonjour; they assume the existence of a special capacity for non-empirical observation, a capacity whose exercise is capable of yielding insights into necessary truths belonging to the domain of reference or semantic values of propositions believed. Bonjour remains open-minded about whether there is a “special” capacity, but for the rest, continues the tradition.

Conceptualists have been quick to spot the weakness of the position. Boghossian claims that “the single most influential consideration against rational insight theories can be quite simply stated: no-one has been able to explain, clearly enough, in what an act of rational insight could intelligibly consist” (Boghossian 2003, 230). He points to Benacerraf’s dilemma as an important problem for the theories. Wright stresses a different problem, having to do with the closeness of understanding and inferring.¹² Taking the example of the Modus Ponens (*MPP*) rule he argues in the following way. Intuition, he says, has

to be capable of going to work in the context of an antecedent understanding of the conditional and an open-mindedness about the status of *MPP*—just as perception can go to work in the context of an understanding of the proposition that I have left my keys in the garage and an open-mindedness about the truth-value

¹² It has been anticipated by B. Stroud in his reaction to Achilles and the Tortoise problem (1979).

of that claim. The point is, however, that there is no such possible context. It is constitutive of an understanding of the conditional to acknowledge, at least implicitly in one's practice, the rule of *MPP*. So an understanding of the conditional cannot coherently be supposed to provide the material—for an intuitive recognition that the rule is sound. If it could, there ought to be such a thing as understanding the conditional perfectly yet—because of a failure of one's intuitive faculty rather than one's faculty of comprehension—failing to be arrested by the validity of the rule. That there is no such possibility means that here there is no work for intuition to do—there is no space for it to work in. (Wright 2004b, 168)

I shall not go into evaluating these arguments, but will pass directly to the conceptualist alternative and its justification strategies.

2 Conceptualism: constitution, compellingness and obviousness

Let me now put epistemological conceptualism on the map of main contemporary positions. It rejects both Quinean and Putnamian appeals to empirical verification, and the contrasting reliance upon a special, logical intuition sensitive to the domain of reference-semantic value. Instead, it turns to the domain of sense, and attempts to derive justification (warrant, entitlement) from our understanding of logical concepts. The view has been proposed, in a line that owes a lot to the work of M. Dummett, by C. Peacocke, P. Boghossian, B. Hale and C. Wright, and in a very different guise, by P. Horwich. The conceptualist approach concentrates upon our grasp of relevant concepts (e.g. in the case of logic, those of *CONJUNCTION* or *IMPLICATION*) and upon their alleged *a priori* connections (e.g. those embodied in their intro- and elimination rules), and attempts to account for *a priori* knowledge in terms of the grasp. It typically attempts to preserve realistically factual character or substantiality of *a priori* knowledge (the exception being Wright and to some extent Hale). At the same time typical recent conceptualists subscribe to a version of naturalism, albeit a somewhat weak one, and attempt to neutralize the worries

traditionally connected to causal or causal-like explanations. The idea is that the mere possession of concepts, no matter how it has been arrived at, provides the thinker with substantial (factual) knowledge about the items concepts refer to. The conceptualist development has been quite impressive, combining, in an innovative and interesting way, elements from two groups of candidate justification (or warrant, or entitlement). The first, the properly concept-focused justification strategy derives from the alleged role of central logical propositions (principles) in constituting logical concepts or meanings. On the negative side, this strategy stresses the alleged superfluousness of empirical explanation of our having logical knowledge and of its reliability. In the present paper Peacocke will figure as its prominent proponent (I shall discuss Boghossian in a paper meant to be the sequel of the present one). The second group comprises considerations having to do with cognitive features of logic external to its narrow formal structure. In this group we find obviousness and compellingness of principles and of simple inferences they support, and the general indispensability of logic. It is, in the words of C. Wright, “indispensable in rational enquiry and in deliberation” (Wright 2004b, 164).¹³ These three groups of justifiers or warrant providers are the main topic of the present paper.

Before passing to the normative issues, let me note a descriptive-explanatory lacuna in the conceptualist proposal. Assume it is agreed (with the conceptualist) that the naïve cognizer’s, say Dorothy’s, ability to reason about situations characterized by conjunctive propositions, is connected to her having the concept “&,” i.e. CONJUNCTION. This still leaves the question open of the *direction of explanation*: does her ability to grasp such situations *derive* from her conceptual ability, or is it the other way around, namely her conceptual ability *deriving* from a more fundamental ability to combine items in the domain of reference (for instance, regarding the broken leg situation in combination with the laceration situation). The conceptualist has to offer an argument for supporting the first direction of explanation, rather than the second. I am not aware of any such argument in the offing.

¹³ Wright has presented his misgivings about both details of Boghossian’s project and about its firm commitment to realism in Hale and Wright 2000. But he endorses his general conceptualist strategy of rule-circular justification.

2.1 The constitutive justification: is the circle wide enough?

Let me remind the reader of Peacocke's conceptualist account, and raise my doubts as I proceed. First, and very briefly (all too briefly, I am afraid), the constitutive justification, Peacocke's main contribution to the subject. He famously starts by claiming that the having of logical concepts is closely tied to (if not identical to) respecting of standard inference patterns that govern the use of the constant signifying these concepts. These patterns, which individuate logical concepts are, of course, codifiable by principles of inference. This is connected with an important cognitive-epistemic element: the person who has the concept, must find instances of principles of inference "primitively obvious" or "primitively compelling." To return to our Dorothy example, she has to find compelling the simple reasoning about the problems her dog is having: it must be impossible for her to conceive of her dog having a broken leg and a laceration without having a broken leg. Obviousness and compellingness, Peacocke claims, go well with constitutiveness, so the strategy of appealing to the former has to be combined with the appeal to the latter. Here is the combined strategy, in a nutshell illustrated by the simplest case, the conjunction.

The typical inference pattern of conjunction, encompassing its introduction and elimination rules is critical for it. Being able to follow the rules, and finding them compelling just *is* to possess the concept "&"(CONJUNCTION). Now, the sense, captured by rules, determines reference or semantic value, in this case truth-value. Nothing else counts. So, once Dorothy has the concept "&," she will find the transition from "the dog has a broken leg" plus "the dog has a laceration" to "the dog has a broken leg & the dog has a laceration" unproblematic and compelling. Consider now the semantic values, i.e. truth or falsity of these sentences. The truth-table for "&" tells us, in complete harmony with the introduction rule, that the latter sentence is true if the two former ones are. So, there is a simple determination principle, going from the sense of "&" to the truth-value (reference) of the composed sentence, as captured by the truth-table. Now, being able to follow the rules, and finding them compelling entails believing the truth of the compound sentence if you believe that its constituents are true. This is knowable from the armchair, so to speak: Dorothy should be able to figure out,

by reflection alone, that the rule she is following entails that the dog will have both problems if it has each of them.

Let me put this into Peacocke's more sophisticated terminology. For the given logical concept, i.e. the concept signified by the corresponding logical constants, call it *C*, there is an inference pattern *I* that captures the sense of *C*. *It also partly determines (individuates, constitutes) it* (Peacocke 1992, 805), since concepts are partly individuated and constituted by their possession conditions. Further, the possession conditions for logical constants are given by inference patterns governing their correct use. In general, having *C* consists in standing in some specifiable relation *R* to standard inference patterns *I* (which is the main internal conceptual consideration); *R* consists in finding instances of the pattern *I* primitively compelling and obvious or "irresistible" (the cognitive-epistemic condition). Given all this, there will be logical concepts associated with principles of inference such that the truth-value of premises in inference together with the pattern itself determine the semantic value, i.e. the truth-value, of conclusion. Like Dorothy in our example, the thinker who has *C* (and stands in *R* to *I*) can come to believe (and to know) *a priori* that the just performed steps are correct, and she can know this *a priori*. In particular, if some combination of inference patterns is valid/truth-preserving, and sufficiently simple to be grasped by the thinker, she can also come to believe (and to know) *a priori* that this is so.

We are now ready for complications. The account is so simple as to appear blatantly circular. The circle is too narrow: the sense determines the reference, so once Dorothy has grasped the sense, she seems to be automatically justified in believing the items in the referential domain, i.e. the domain of semantic value (true/false). Peacocke wants to avoid this, and introduces a crucial complication. The qualification "partly" in the italicized sentence above plays an important role: it is meant to block the objection that the strategy is blatantly circular. The secret is that relations in the domain of semantic value have a kind of backward-looking effect to the domain of sense. The truth-table for "&" not only fits nicely the introduction and elimination rules for "&"; its availability so to speak makes such rules legitimate. Without this legitimation from the truth-table, the inference rules determining power would lack ultimate justification. This is the secret of the key qualification "partly." In Peacocke's words, the relations in the domain of

semantic value in a way “ratify” the sense, and the sense, in the usual Fregean way, fixes the relations. There is thus a kind of reciprocity: the rules determine the truth-table for “&,” but the truth-table legitimizes the rules. Peacocke aptly calls the conception advocated “the reciprocal conception.”¹⁴ In his view, it answers the question about the relation between justification and the individuation of sense by saying that neither is explanatorily prior to the other.¹⁵

¹⁴ Here is how Peacocke situates his conception in the space of alternatives:

The reciprocal conception thus steers a middle course between two alternatives outlined by Dummett:

(1) “The meanings of our assertoric sentences generally, and of the logical constants in particular, are given us in such a way that the forms of deductive inference we admit as valid can be exhibited as faithful to, and licensed by, those meanings and involve no modification of them. In this case, these principles of inference will indeed be capable of justification, possibly together with other principles we have failed, but are entitled, to acknowledge.

(2) Our principles of inference admit no justification, because they are not faithful to the meanings of our statements as antecedently given, but instead serve to determine the meanings of our logical constants, and, in part of sentences not containing them.

If we are to command a clear view of the working of our language, we have to decide between these two alternatives and then to flesh out the one we have chosen.” (Dummett 1991, 194–195)

The reciprocal conception combines elements of each of the alternatives between which Dummett says we have to choose. On the reciprocal conception, there is certainly an element of faithfulness to a semantic value; and this is an element in Dummett’s first alternative. But the reciprocal conception will certainly also treat certain principles, including some inferential principles in the case of the logical constants, as serving to determine, in part, the meanings of sentences containing the expressions in question. The relevant meaning-determining principles will be those mentioned in the understanding conditions for the expressions in question; and this is an element in Dummett’s second alternative. (Peacocke 1992, 805)

¹⁵ Let me briefly document this short summary. Peacocke’s early and groundbreaking lecture on understanding logical constants (1987) offers a good beginning. It starts with finding instances of inference principles compelling, our step (B). Taking conjunction (“&”) as his example of a constant Peacocke puts forward two claims:

(*Claim 1*) For this logical constant, there are principles containing it whose instances are found primitively obvious by someone who understands it ...

(*Claim 2*) For this logical constant, finding instances of these principles primitively obvious is at least partially constitutive of understanding it (of “grasping the sense it expresses,” in the classical terminology). (Peacocke 1987, 155)

In his later work (most recently 2004, 2005), Peacocke develops these (rather early) ideas, generalizing the notion of inferential principle associated with a concept. Take for any given concept a true statement of what it is to possess it. Such a statement captures the “possession condition” of the concept. The account of how reference or semantic value is determined is now presented as a theory connecting the account of possession conditions for concepts with the determination of the semantic values for those concepts, the “Determination Theory.” The two accounts together make up “a metasemantic account.” Its starting point is the following. First, sense together with the way the world is determines reference. Next comes the important conditional: “If sense together with the way the world is determines

The principles are captured by introduction rule (&I) that allows one to pass from A, B to $A \& B$, and by two elimination rules (&E1) and (&E2) allowing one to pass from $A \& B$ to A , by (&E1), and equally from $A \& B$ to B , of course by (&E2). *Claim 2* turns to instances of the principles. The idea is that “no one understands ‘&’ unless he finds these instances primitively obvious, and that the explanation of this fact is to be traced to what is involved in understanding conjunction” (1987, 155). Conjunction is very simple, and other constants are typically more complicated, so their principles might not be equally obvious, but Peacocke proposes ways to circumvent the problem (since we agree that his proposal works, we shall accept that the problem can be thus circumvented).

Now what is the status of *Claim 2*? Peacocke mentions a legitimate worry, that it might just “illegitimately elevate what is just a psychological generalization about subjects who understand conjunction to a purported necessity.” He answers by denying that *Claim 2* is an empirical psychological claim:

Claim 2 is a claim about a constant *with a given sense*. If it is true at all, *Claim 2* is necessary because it is a claim about what contributes to the individuation of a given sense. It follows from conditions governing the notion of sense that principles which must be found primitively obvious by anyone who grasps a sense also contribute to the individuation of that sense. (156)

His argument, Fregean in origin, relies on an epistemic notion of sense: Fregean sense is individuated by considerations of informativeness, so if a thinker can doubt that the semantic value of one expression is the same as of the other, then the senses of the two are different. And he assumes that sense is given by what is primitively obvious to the cognizer. (I note in passing that in his later work, for instance in *Being Known* [1999], this epistemic principle of individuation is combined with a quite Platonic ontology of sense.) Next comes the link with truth-value:

(*Claim 3*) What makes a particular function the semantic value of “&” is that it is that function which, applied to the semantic values of the expressions on which the conjunction operates, ensures that the principles instances of which are found primitively obvious are indeed genuinely truth-preserving. (157)

reference, and possession conditions individuate senses, possession conditions together with the way the world is must equally determine reference. More fundamentally ... if a sense is individuated by the condition for something to be its reference we must show how a possession condition for a concept *C* fixes a condition for something to be the reference of *C*” (Peacocke 2004, 171, reiterated in 2005, 752). Now, how does the reciprocal conception assure apriority? Peacocke’s preferred account of apriority, which is generated by claims and principles listed, is his metasemantic account. Here is his argument from

Peacocke’s explanation is exactly what one would expect given soundness of propositional logic: “If the rule (&I) is to preserve truth, then *A*&*B* must be true when *A* is true and *B* is true. Similarly, *A*&*B* must be false when either *A* is false or *B* is false; if it were not, then either (&E1) or (&E2) would fail to preserve truth” (Ibid.). But this cannot be a whole story. An obvious problem with it is that it moves in a very small circle: if rules of inference completely determine truth-values, we cannot appeal to intuitive correctness of the truth-table to support rules. Peacocke therefore insists on two additional points: rules of inference only *partly* fix the semantic value, and, consequently, neither of the two, rules and semantics, has priority over the other. In a later paper, “Sense and Justification” he reiterates that “the semantic value of a logical constant is fixed as that function which makes truth-preserving those transitions involving the constant which one who understands the constant must find primitively (underderivatively) compelling” (Peacocke 1992, 803). But, he adds the following:

This approach still leaves room for justification, for when a proposed logical constant does have a sense, there will be a semantic value for which its distinctive methods of inference are correct. This kind of justification is not vacuous, since there are some “roles” for which no truth-preserving semantic value can be given, and those for ‘*tonk*’ are amongst them. On this approach then, for a constant of the propositional calculus, it is a substantive requirement that it have a genuine truth-table, and we inherit this attraction of the purely extensional approach to Frege’s “stipulations.” But since the truth-table does not uniquely fix the sense—two different sets of rules may determine the same truth-table—the obstacle to the purely extensional treatment is removed. (803)

As mentioned, he calls this “the reciprocal conception”: neither justification nor the individuation of sense is explanatorily prior to the other:

In one direction the semantic value of an expression with a given sense is fixed through certain principles which must be accepted by anyone who fully understands the expression. In the other direction, those principles themselves are ratified as correct by the semantic values so determined. (804)

So the principles are not viciously self-justifying, for the role of semantic values is essential in justifying them. But they do have to a certain extent a “self-involving justification”: The justification of those principles proceeds via a semantics for the expressions, where the semantics is the right one because it validates those principles (805).

the same sources. A content p is outright *a priori* if the possession conditions for the concepts comprising p together with the Determination Theory jointly guarantee the truth of p . An outright way of coming to know p is an *a priori* way if the possession conditions for the concepts in p together with the Determination Theory jointly guarantee that use of that way leads to a true belief about whether p is the case. Similarly, a transition from one set of contents to a given content is an *a priori* transition if the possession conditions for the contents involved together with the Determination Theory jointly guarantee that the transition is truth-preserving.¹⁶ This is the end of our brief, all too brief, summary.

Time for some criticism. One should keep in mind that the strategy proposed by Peacocke is a combined one, while discussing its component substrategies. We start by briefly addressing the issues of constitutivity. The point that has attracted most attention is Peacocke's view that having the relevant logical concept entails accepting the associated principles-rules, and finding them compelling. Let me just mention two criticisms of this view with which I agree. Tyler Burge (2003) has been arguing that conceptual understanding might fall short of requirements Peacocke and other conceptualists impose upon it. T. Williamson (2006) has been recently discussing and criticizing views linking understanding of alleged conceptual truths with assent and knowledge. His example concerns very simple propositions, like "Every vixen is a vixen." He wonders about claims like the following: Necessarily, whoever grasps the thought that every vixen is a vixen believes (knows) that every vixen is a vixen. Necessarily, whoever understands the sentence 'Every vixen is a vixen' assents to it (recognizes it as true). His imaginary counterexamples concern two native speakers of English, Peter and Stephen. Peter's first step in evaluating these claims is

¹⁶ The old "principles" for "&" are now called "possession condition for the concept of conjunction." We know that cognizers must find the transition from $A \& B$ to A compelling by itself. A plausible Determination Theory will entail that semantic values are assigned to concepts in such a way as to make truth-preserving any transitions which, according to the possession conditions for a concept, must be found compelling. Again, if the rule (&I), contained within possession condition, is to preserve truth, then $A \& B$ must be true when A is true and B is true, and so on. "It is thus a consequence of the possession condition for conjunction, together with the Determination Theory, that when $A \& B$ is true, A is true. That, according to the metasemantic theory, is why the transition is *a priori*" (Ibid.).

to notice that they seem to presuppose the existence of vixen. Then, Peter comes to the considered view that the presupposition is a logical entailment and that universal quantification is existentially committing. Peter also has the weird belief that there are no vixens. So, he seems to be a counterexample to the claims used to illustrate the link between understanding and thinking the thought on one hand, and knowledge and belief on the other. The other hero is Stephen who produces counterexamples to the claims, derived from considerations of vagueness. Beliefs about logic and matters of fact can block the passage from understanding to believing, accepting as true and knowing. The criticisms mentioned show, in my view, that the constitution-strategy is far from compelling. I would like to connect them with a wider question, concerning two topics which have not been so prominent in the debate.

The first is the issue of the circle of justification based on concept-constitutiveness. There are several elements of circularity threatening Peacocke's proposal. Two are particularly worrying. First, rule circularity: reasoning about justification is, of course, using the rules that are to be justified. Peacocke has nothing to say about it, so I will also leave it aside here.¹⁷ Second, and more interesting, the "reciprocity" between possession and determination, between rules of inference and rules concerning truth-values (and semantic values in general). Although Peacocke sounds very enthusiastic about this "reciprocal conception" allegedly avoiding the pitfalls of circularity, it is clear that his proposed strategy works within a single domain, justifying like by like. At the professional level, semantic truth-tables are part of logic as much as deduction rules are. Reliance upon truth-tables is a matter of the same quite narrow capacity as the confidence in deduction rules. We know it, because we aim at connecting the rules and the tables when looking for soundness proof(s), and we do it by using the very same logic we are investigating. More interestingly, the same immanent character is manifest in ordinary or naïve reasoning. Our conceptualist might propose that thinker's rule-and-principle-focused interests and capacities are more abstract (or, alternatively, more introverted, turned to mental representations and the like), whereas thinker's semantic

¹⁷ Of course, this issue is discussed in detail by Boghossian, in his version of conceptualism. However, he applies it to a different step in the conceptualist program (as noticed by my student A. Butković in the discussion; I thank her for reminding me).

interests and capacities are more world-oriented (or at least model-oriented).¹⁸ However, his correct insistence upon finding *instances* of rule-and-principle compelling blocks this line for him. Finding an instance of a rule compelling means finding a particular case compelling, and here the distinction more or less breaks down. To see this, we shall return to our dog owner. From the conjunction about the broken leg and laceration she is able to derive the conclusion about the broken leg. She is also capable of finding out that it is (or at least seems) impossible that the conjunctive state of affairs obtains, without each of its components obtaining. However, if this is testimony to her “appreciating in the right way that the truth of the premises guarantees the truth of the conclusion,” the way Peacocke represents the matter, then her two abilities, both the derivation-focused and the truth-appreciation focused ability, are components of the same ability, working within the same larger domain. To use the terminology that Peacocke himself does not use, the reciprocal conception brings the two patterns, inferential and semantic, and intuitions about them into reflective equilibrium, but the equilibrium is a (very) narrow one.

Here is an invidious analogy with sensory modalities. When my spectacles are not around, I might be unsure about a shape of something in front of me, and I can help myself by feeling the object, in order to make sure that the visual information I am receiving is reliable. This is a very practical method for a practical context, but of little use in theoretical debate about reliability of sense-perception. One component of perceptual ability is tested by its conformity with another, within the same perceptual domain. But isn’t this similar to what happens when we justify our trust in deduction by looking at truth-table and *vice versa*. Is not reciprocity just a nice-sounding name for circularity? Call this “single-domain problem.”

The literature on coherentism has familiarized us with a recipe: if a circle of justification is too narrow, try to widen it. Why is wide circularity better, some of my colleagues wonder. Well, for one, a wider data-base answers more objections, as Zsolt Novák put it in a discussion. Of course, if you are a foundationalist, no amount of coherence will by itself justify a belief for you; but there is a widespread

¹⁸ Thanks go to my colleague and friend Nenad Smokrović, who raised the possibility in a discussion.

agreement in the literature that a wide circle of coherent beliefs is in good shape to be justified. Now, how could the circle under discussion be widened? An obvious proposal is to widen the reflective equilibrium. At a reflective level we might introduce the worldly success of logic, indicating its actual reliability, as an empirical reflective component of justification. I am saying “a component,” since there is no reason not to accept others, of a more *a priori* nature. To see this, let us stay just for a moment with the world-oriented scenario characterizing the key inferential moves and intuitions of naïve thinkers. Remember that Peacocke himself said that “in respecting the norms of embedded rationality the thinker is also respecting those that are made available precisely because he is embedded in his world in a certain way,” and moreover has referred to “states which are individuated by the ways they are embedded when the thinker is properly connected to the world” (Peacocke 2004, 178). So, back to Dorothy, the dog owner, embedded in her world. We, epistemologists wonder what makes her justified not only in making her inferential transition, but in trusting them when it comes to imagining. Let me explain. She cannot imagine the conjunctive state of affairs obtaining without its component substate obtaining. Why is she justified in taking her inability to imagine otherwise as a sign that the problematic situation will not obtain, that it is *objectively* impossible? Once we concentrate upon the semantic values and worldly configurations, instead of formulae and senses, we might understand a holistic temptation: the reasoning of this kind has proved to be massively *successful*, so why could not a reflective dog-owner appeal to this past success as *one reason* for her trust? She is not thereby becoming a fanatical Quinean. She can, first, retain the obviousness and compellingness as her immediate reason. But if questioned further, about credential of such immediate obviousness, she might reply that it has not deceived her ever. It is, to use a fashionable turn of phrase, a virtuous trait. Further, these more holistic, success-involving considerations might be taken in the usual way to indicate objective reliability of one’s methods. This is the way realism functions in science and in commonsense. A good, probably the best, explanation of success is truth(-likeness) and reliability. Such a wide justification would be a higher, more reflective one, strengthening the credentials of immediate obviousness, and not necessarily replacing it with empirical considerations.

But why worry, an advocate of narrow justification-through-constitution might respond? Is not the narrow circle sufficient? No, it is not. The single domain problem shows its teeth when we address the most general semantic-metaphysical question: How does the thinker's conceptual repertoire constrain reality it refers to? How does determination work? This brings us to the issues about realism and anti-realism:

In his "Understanding Logical Constants," Peacocke addresses the issue following Wittgenstein's lead. Logic is not "hyper-physics." And logical "laws" there in the world do not make or cause things to be this or that way. Things go the other way around: it is the logical rules constitutive of concepts that create new facts:

So in these primitive cases, it would be better to say that the principle's validity consists in part (though of course not wholly) in an impression that it is valid; and this seems incompatible with its validity causally explaining the impression. The failure of the claim that validity causally explains impressions of validity gives a limited sense in which these primitively obvious principles do not hold independently of our impression that they hold. This limited sense is analogous to that in which, on contemporary accounts of secondary qualities, the redness of a perceived object does not hold independently of a normal perceiver's experience of it as red in normal circumstances. (Peacocke 1987, 79)¹⁹

The crucial idea is that "the principle's validity consists in part (though of course not wholly) in an impression that it is valid."

¹⁹ Let me, for the sake of documenting the view, quote the immediately preceding lines:

The fact that something is square can causally explain a subject's experience of it as square. Now take a primitive logical principle, a principle such that to find it primitively obvious is partially constitutive of understanding the expressions it contains. On the present account, one ought not to try to explain causally the fact that a thinker finds it primitively obvious that a certain principle is valid by citing the fact that is valid. The principle's being valid consists in its being truth-preserving under all relevant assignments; it is truth-preserving under all relevant assignments in part because of the semantic values given to the logical constants it contains; and in turn these constants receive their semantic values in part because the thinker finds it primitively obvious that the given conclusion follows from the premises. (Peacocke 1987, 79)

Logical properties are not like primary qualities, but more like secondary ones! A touch of anti-realism seems to be congenial to conceptualism, since conceptualism attempts to replace causation by conceptual determination, going from the domain of sense to the domain of reference, i.e. objects and facts. But such a determination seems to introduce a strong element of anti-realism. Call this “order of determination problem.” Would a distinction between epistemic and mental help? Maybe it would, but Peacocke’s proposal featuring “an impression” of validity is definitively on the mental side of the divide, as is the comparison with red.

To illustrate the problem further, let me quote other conceptualists like Bob Hale and Crispin Wright, who recommend that we embrace quietism, give up the idea that the true *a priori* proposition mirrors an independent reality and “acquiesce in the conception of general kinds of things the world contains which informs the way we think and talk, and is disclosed in our best effort to disclose the truth” (Hale and Wright 2002, 122). The cure they propose is quite radical: their conceptualist account leaves room “for the capacity of such explanations to *invent* meanings” (2000, 295). They claim that it avoids Benacerraf’s dilemma; for my part, it looks more like embracing its anti-realist horn, than like a neutral, truly quietist way out. (Hale himself, in his book *Abstract Objects* (1987), and in his non-collaborative papers, seems to be closer to realism, more precisely to Platonism without causal contact with human minds.)

Interestingly, as Peacocke is generally sympathetic to realism, he never reiterated his secondary quality claim (but also never performed an official recantation, at least as I was able to check). Recently he has made a further realist step, trying to build elements from the referential option into his conceptualism. Here is a crucial passage:

The conditions which individuate the entity in question (the set, color, number, shape) actually enter the possession condition for certain canonical concepts of these entities. As one could say, in these cases, the concept is individuated by what individuates the object. The implicit conception detailed above which underlies mastery of the notion of a whole number already exemplifies this phenomenon. The content of that conception specifies what it is to be a whole number. (Peacocke 2005, 757)

However, the question remains. If there is such a close link between individuation of concepts and of objects, and if the order of determination is now reverse, and it is the latter that dictates the former, we seem to have a new problem. How does the thinker access such an object-determined concept? When reading the earlier account, one assumes that concept-possession is not very problematic: human beings somehow arrive at the relevant rules (or have them hardwired in their minds/brains), and the rules constitute the concept. But, if *logical reality* is already built into logical concepts and into the rules, then we need a substantial story. Merely appealing to the principle that sense determines reference does not work. In fact, the problem was already there in the original account. Remember Peacocke's claim that finding instances of constitutive principles primitively obvious is at least partially constitutive of understanding the logical constant (Peacocke 1987, 155). The claim is not psychologistic in a bad sense, he argued, since "it is a claim about what contributes to the individuation of a given sense." And "principles which must be found primitively obvious by anyone who grasps a sense also contribute to the individuation of that sense" (156). However, if the connection between the two is as tight as claimed, we do have a mystery. Suppose that the fully objective sense is there, in the Platonic domain of senses, equipped with principles that contribute to its individuation. Then we have a version of Benacerraf's problem for it: how is such a Platonic entity grasped by the thinker? Even if we accept the optimistic assumption that one who grasps the sense should somehow find principles obvious, this only makes things worse: what kind of harmony between the objective sense-cum-principles and human cognitive capacities accounts for actual experience(s) of primitive obviousness? If, on the other hand, we take the individuation of sense itself to be dependent upon human "ratification" of partly constitutive principles, partly "invented" by us humans, then the objectivity is weakened, and the serious realism is gone. (If we "depsychologize" the ratification, we are back to Platonism.) C. Wright wins and Peacocke loses. Some authors following Kant claim that there is a third option, the "transcendentalist" one, which merits a longer discussion. I am afraid that it collapses into mind-dependence, but shall not discuss it here, since the present-day conceptualists don't even mention it.

Would a wide reflective equilibrium help the realist, by bringing empirical success into play? It seems to me that it would, and

moreover, that Peacocke should himself acknowledge it. Consider the quotation from the section we have already looked at above, talking, in the context of logical knowledge, about embedded cognizer and the norms being “made available precisely because he is embedded in his world in a certain way” and of “states which are individuated by the ways they are embedded when the thinker is properly connected to the world” (2004, 178). The view of the function of norms is placed squarely within the world-oriented approach. Now, if it is the features of the world itself that enter the concepts our thinker has available to herself, what is the rationale of excluding in advance the empirical information about the world from the pool of available justifiers? As long as we have dealt with purely conceptual considerations, the exclusion seemed reasonable; but if the very cognitive states of the thinker relevant for logical beliefs and knowledge are individuated by the ways they are embedded when the thinker is properly connected to the world, isn’t it dogmatic to refuse the idea that the empirical information, say about proper connectedness, might play a justificatory role after all?

2.2 Cognitive-epistemic and explanatory considerations

This brings us immediately to the already mentioned cognitive-epistemic ingredient of logical knowledge, namely compellingness which is not blind, but luminous. And this component cries out for an explanation, as we have just seen. But let us proceed more cautiously, and first consider the compellingness itself. The mere mention of “primitive compulsion” to θ invokes above all the impossibility to do (in the relevant cases to infer, imagine or conceive) otherwise than to θ . Such an impossibility would solve the justification problem by deploying the “ought implies can” principle, and would offer some guidance to the cognizer and evaluator.²⁰ But it is too blind and mechanical and therefore not congenial to most conceptualists, since they do not think of concepts and grasping of concepts in such mechanical terms. This is why Peacocke needs obviousness or some related kind of “ra-

²⁰ This is, for instance, P. Horwich’s tactic in his paper published in Boghossian and Peacocke 2000.

tional” compulsion.²¹ Elementary logical moves or transitions, as well as intuitions accompanying them, are characterized by luminous understanding, involving an awareness that these moves are necessarily valid. In his earlier work Peacocke spoke of primitive obviousness or primitive compulsion, but in recent work he adds the requirement that thinker should be “rational, from his own point of view” (Peacocke 2004, 178). Obviousness involved in “seeing” that the transition in question is correct is also essential for the satisfaction of the requirement of rationality from the subject’s own point of view: “It [i.e. the satisfaction] invokes the kind of reasons that one ordinarily invokes when justification is required, as in the dialogue ‘Why do you think that?’—‘I see it to be so’” (178).

²¹ Compare Hale 2002. He starts from the rule-circularity worry.

Suppose it is granted that we have an explanation why our basic rules are immune to relevant doubt. To advance from this to the conclusion that claims about their soundness are true, and known to be so, will surely require some reasoning. And it is hard, if not impossible, to see how any reasoning to the purpose could avoid using at least some of those very rules, so that any attempt to close the gap between our intermediate and our final conclusion must be rule-circular. Further, the rule-circularity involved would be of precisely the kind I have claimed to be especially problematic, because it would occur in the context of an attempted explanation of how we can come to know something. So, unless it can be argued that that kind of rule-circularity is, after all, tolerable, we seem to be stymied.

And he almost immediately passes to the impossibility of rational doubt.

A general model of knowledge—of inferential as well as non-inferential knowledge—which would accommodate my suggestion might run roughly as follows. To know that *p* is to have a true belief that *p* and to be entitled to that belief. There are various much discussed ways in which a thinker may be so entitled. She may have a warrant to believe that *p*, in the sense that she has reasons or grounds to believe that *p* which she can articulate. Or she may have acquired her belief that *p* by a reliable method which does not involve her having independently ratifiable grounds or reasons for her belief. Perhaps she can just see or hear or otherwise sensorily detect that *p*, or be introspectively aware that *p*—that is, it may be enough that she be sensorily or introspectively affected in ways that induce in her a belief that *p*, provided that there no special grounds to suspect that she is victim to some perceptual or introspective illusion or malfunction. Alternatively—and to a first, crude, approximation—she may satisfy the entitlement condition by believing something which is impossible—and so impossible for her—rationally to doubt. (2002, 302–303)

Note that no explanation of the impossibility has been offered. Similarly, Boghossian does not even discuss issues connected with psychology of compellingness.

the rationality from the thinker's own point of view of a logical transition consists in his appreciating in the right way that the truth of the premises guarantees the truth of the conclusion, where the truth of the conclusion is conceived in accordance with the thinker's tacit knowledge of the contribution made to truth-conditions by its logical constituents. Under this approach even primitive axioms and inference rules can be rationally accepted.

It also permits the appreciation of the validity of new axioms and rules that do not follow from those previously accepted.

In respecting the norms of embedded rationality the thinker is also respecting those that are made available precisely because he is embedded in his world in a certain way, or is in states which are individuated by the ways they are embedded when the thinker is properly connected to the world. (178)

This is, of course, just a very brief reminder. Before passing to discussion proper, let me reiterate a basic question: What is exactly the object of the luminous, rational compulsion? "Finding certain transitions compelling" is ambiguous between an introverted, thought-focused meta-level reading and an extroverted (world-focused, first-level one). At the meta-level one finds the theorist reflecting about whether an instance of a principle, represented by a formula, would be acceptable. Take the incompatibility of instances of " p and q " with non- p . A logic teacher sees immediately that the formulae are not jointly satisfiable. This is meta-level insight. At the first level the target is different. We have already mentioned the extroverted, world-oriented character of ordinary inferences: it is not thoughts themselves (instance of p -thought and the rest) that appear incompatible, but situations or states of affairs that exemplify them. Take Dorothy the dog owner from our examples. Of course, she finds the situation itself, a-dog-having-a-broken-leg-and-a-deep-laceration-without-having-a-broken-leg inconceivable, not something about the sentence or thought. And this is from the epistemological standpoint the *primary kind of logical incompatibility*. When a naïve thinker begins to reflect, a given simple inferential step might seem obvious to her. But not in the way the logic manual presents it. Unfortunately, the standard talk about concepts and rules suggests more immediately the introverted, meta-level considerations and this suggestion is

phenomenologically incorrect. Of course, it is always possible that the extroverted awareness about situations results from some rule-based source, either hard-wired rules, or implicit representation of them, but in the phenomenology the extroverted awareness comes first. We shall revert to this fact later. The actual account is rich and intricate, a fruit of more than two decades of philosophical reflection.

The classical problem for obviousness-cum-compellingness is the dilemma between a psychological and a normative-epistemological interpretation. Obviousness-cum-compellingness is one of the two: either (a) a psychological-cognitive property of the proposition pointing to a mental state of “seeing” or “being certain,” for which the experiential feel is essential, or (b) a normatively characterized epistemological property, relatively independent of psychological realization, for instance independent of any particular experiential feel of the state, and such as to be characterized in terms of what is sometimes called “self-evident” character: once a cognizer understands the proposition, she is rationally permitted or even obliged to come to believe that it is true, and the like.²² If (a), then the question arises how to pass from psychological, subjective certainty to any kind of objective assurance. For instance, why would mere feeling of compulsion indicate that some links are conceptual as opposed to be merely empirical and contingent?²³ Moreover, (a) is not only clearly compatible with usual naturalistic explanation-attempts, but actually cries for an explanation. If (b), the opposite question arises: how does the thinker recognize the normative property in question? Take the ubiquitous error of the denial of the consequent. Many people find it compelling to pass from “If she likes me, she will go to the movie with me” to “She doesn’t like me, so no movies together.” What is the connection between this psychological fact and the normative realm? Why is the compellingness here something problematic, that has to be either explained away (well, this is just an inductive guess, not really a logical mistake), or justified by special, additional means (the thinker took “if” to introduce

²² Both alternatives have been attributed to Descartes, and defended by contemporary epistemologists. Alternative (a) seems more prominent in Ayers, and perhaps BonJour, alternative (b) in recent work of Jeshion (2001) and Evnine (2001).

²³ It was raised, for instance, by G. Rey a decade ago in his criticism of Peacocke (Rey 1996).

an equivalence, not a conditional)? Why is in other cases psychological compellingness taken to be a good indicator of the presence of the corresponding normative constraint? The reading encapsulated in (b) tells us strictly nothing about these issues. Now, Peacocke, like most conceptualists, takes obviousness-cum-compellingness very, very seriously: it is an essential justifier, and indicator of the normative status of logical propositions, as *Claims 1–3*, quoted in note 15 clearly document. He, and they, therefore, owe their readers an account of what this property really amounts to. We shall return to this debt presently, in the context of more radical skeptical doubts about obviousness. Before this, let us consider the related issue of explanation, or the lack thereof.

How is the thinker's access to logical matters to be explained? In the conceptualist view determination can replace causation. Conceptualists all rely upon the alleged ability of sense to determine reference in an *a priori* knowable manner. Peacocke is one of the pioneers of this program; we have seen how in his view the sense of logical constants is linked to inferential practice and to luminous understanding, how the sense enters reciprocal relation of determination-ratification with the domain of semantic value and how all these components offer an *a priori* route to the domain of logic. Peacocke stresses that this "metase-mantic" account is sufficient, and that we need no causal account of how we come to know *a priori* and modal truths. That there are no "special faculties" involved in our coming to know them, and, even more importantly, any causal explanation is bound to be irrelevant.

The fact that the truth that *p* explains one's belief that *p*, and perhaps by some special causal route, involving some postulated special faculty, fails to imply a crucial feature of some cases of the *a priori*, which is that *p* will hold in the actual world, whichever world is the actual world. In short, any faculty conceived on a quasi-causal model, far from helping to explain the phenomena of rational intuition and *a priori* knowledge, is actually incompatible with the nature of the phenomena to be explained. (Peacocke 2004, 167)

The crucial feature that resists causal explanation is a close cousin of traditional necessity, which was famously similarly taken by Kant and

his followers to preclude causal explanation. Here, Peacocke suggests, “we are well rid of any attempt at a causal epistemology. Attempts to develop the epistemology of the *a priori* or the modal in causal terms can only encourage the view that defenders of the *a priori* and of necessity must be committed to unacceptably non-naturalistic conceptions. One motivation for that charge is removed if our epistemology of these two notions is not causal” (168). The “metasemantic” account is sufficient, and we need no causal account of how we come to know *a priori* and modal truths.²⁴

In order to judge the prospects of such an anti-explanationist strategy, let us first remind ourselves that the inferential activity to be understood is the extroverted, semantic value oriented and world-focused way. The obviousness or luminosity is tied to the worldly situation represented—or to the process of representing. This extroversion might be the key to luminosity: it is our ability to think of semantic values that gives us luminous understanding. However, once we take the extroverted view and worldly situations as the primary target, the conceptualist account already seems a bit thin: it is in a sense true that the dog owner might represent the transition from composite situation (broken leg and laceration) to the complement of one component (no laceration) impossible in virtue of having the concept CONJUNCTION. But the example immediately suggests more, namely the ability to aggregate, to think of states of affairs or situations as *taken together*; possibly, her mastery of “&” might derive from this more primitive ability. All this *leaves open* the relation between having concept CONJUNCTION, mastering “&” and being able to think of states of affairs taken together. Maybe she can understand the sentence in virtue of her more primitive ability to aggregate in thought. And what about negation and complement?²⁵ What abilities are involved in conceiving of

²⁴ Here is more:

Simply adding that *p* has that further property [i.e. to be true in the actual world, whichever world is the actual world] is to take for granted the contentually *a priori*, rather than giving an explanation of it. If it is specified that there is, in addition to the alleged causal interaction, some further feature of a way in which someone can come to believe *p* that ensures that *p* will hold whichever world is the actual world, this further feature is then doing all the work in explaining the *a priori* status. (Ibid.)

²⁵ For a fine discussion of these issues see Evnine 2001.

the truth conditions of a negative statement? It is quite incredible that philosophers would just give up on explanation of these abilities, that they would refuse to inquire how it could in principle happen that humans have (developed) these abilities, how come that they are tracking their targets so successfully. But this is precisely the program of conceptualism.

To get the feeling of how problematic this strategy is, compare Peacocke's treatment of justification of perception appealing to proper functioning of our cognitive faculties, and involving at least some kind of considerations of the causal structure underlying it. One would expect some analogy between first order rational and intuitional knowledge with a perceptual one. Interestingly, Peacocke is very friendly to causal, even evolutionary Darwinist account, when it comes to perception. Let me summarize. Any foundationalist epistemology has to account for transitions from a more foundational level to upper levels built upon it, e.g. from being-appeared-to redly to "This looks red" or directly to "This is red." Let's call the transition in perceptual case "appearance-reality transition." When is a perceiver entitled to such a transition, and in general when is a thinker entitled to make a transition in thought? Peacocke's rationalism offers an answer in terms of epistemological principles, concerning such entitlements. According to his Principle I, also called The Special Truth-Conduciveness Thesis, a fundamental part of what makes a transition T "one to which a thinker is entitled only if the transition tends to lead to true judgments (or, in case the transition relies on premises, tends to do so when its premises are true), in the way characteristic of rational transitions" (Peacocke 2002, 385). But which way is so characteristic? The answer is given by Principle II, The Rationalist Dependence Thesis: "The rational truth-conduciveness of any given transition to which a thinker is entitled is to be philosophically explained in terms of the nature of the intentional contents and states involved in the transition" (390). What is the epistemological status of such content and state-based entitlement? Principle III, The Generalized Rationalist Thesis offers a simple answer: "all instances of the entitlement relation, both absolute and relative, are *a priori*" (394). In the case of perception they have to be such, on the pain of regress, Peacocke argues. Classical rationalist foundationalists, prominently Chisholm, have been putting forward similar views. And they also held that the ultimate rationale for

accepting principles governing the main types of transition is that they are evident and indispensable. Peacocke has a different proposal for the case of perception. There is rationale for perceivers being entitled to appearance-reality transition, and it is a version of inference to the best explanation. The claim that “This is red” gives, under favorable circumstances, the simplest explanation of why this appears red. And the more detailed story will appeal to evolutionary considerations as contributing to simplicity and the high quality of explanation. In Peacocke’s view

the Darwinian legacy is of significance even in the relatively *a priori* domain of theories about the normative notion of entitlement. This significance does not result from a confusing of the normative and the descriptive. Rather, the claim is that a proper philosophical explanation of certain truths about the normative—the entitlement relation—must be accounted for by the special explanatory status of Darwinian mechanisms. What has been important for the argument is not the empirical truth of Darwinian hypotheses but the special, complexity-reducing status of explanations by some natural-selection mechanism. (Peacocke 2004, 108)

So, causal explanation can be very relevant for epistemological account, Peacocke proposes. But not so in the case of epistemology of *a priori* domain. The proposed asymmetry between perceptual and *a priori* case is quite worrying. Why does not the more puzzling kind of candidate knowledge require deeper explanation? Classical apriorists, culminating with Kant, thought that the question of how we could possibly gain *a priori* knowledge is the prime question of epistemology as a whole. Can it be really demoted to a non-issue, simply by appeal to our possession of concepts? And we have noted that Peacocke himself has claimed, for instance in his paper “Explaining the *A Priori*,” that in some central cases of candidate *a priori* knowledge, like arithmetic and geometry, the theory has to appeal to sensitivity to constitutive properties in the domain of objects that seems to go beyond mere concept possession, and rather than being explained by our concept-possession it in its turn explains the latter (Peacocke 2000, 274). Where does this sensitivity come from? Do we not need an account of

it, and would not that account enjoy a degree of priority in relation to the metasemantic one?

In fact, we do have some building material out of which one might try to construct an explanation of reliability of human logical abilities. It is the view of logic as basic for our inferential capacity that constitutes the common ground between warring factions on the contemporary scene, the common ground of many naturalists and anti-naturalists from Quine, through Dennett and Davidson to, say, Bonjour (it has perhaps been obscured through the visibility of Chermiak-Stich program and the anti-logician tendencies in some psychological research, that have created an association between naturalism and radical anti-logicism). This justificatory basicness is neutral, i.e. compatible with ultimate apriorism and ultimate aposteriorism (e.g. with logical intuitions being ultimately, but not *prima facie* justified by our total theory). So, assume that logic is universal, i.e. minimal rationality very probably involves practical understanding of logical constants. But then, any process that results in there being minimally rational creatures, will also very probably result in there being minimally logical creatures. Applied to evolution, if evolution produces rational creatures, then it very probably produces minimally logical creatures. Given evolutionary naturalism, minimal logical capacities are to be expected. But why would evolution bother to produce rational creatures? (Notice that nobody would seriously question causal explanation on the ground that the emergence of life itself is relatively improbable, or that evolution of the nervous system is antecedently quite improbable. The challenge is more geared to the alleged possibilities of evolution of intelligent creatures similar to ourselves, but devoid of logical rules.) Minimal practical rationality brings practical advantages, so the antecedent itself is not so wildly improbable as anti-naturalist critics (from Plantinga to Nagel) would have it. Although the evolution of rational creatures is in itself “radically contingent,” the fact that once rational creatures are on their way, minimally logical creatures are on their way is not radically contingent. We have no reason to doubt our own rational capacities and basic logical intuitions on the grounds of evolutionary theory. (Evolutionary theory does not undermine itself.) There is no reason not to attempt to dig deeper in this direction. So much about explanation.

Let me now return to the issues of compellingness-cum-obviousness and connect them with the third justifier prominent in the

conceptualist program, indispensability. A classical problem in the epistemology of logic is the Achilles and the Tortoise issue: Basic inferential steps are either learned from explicit rules, or performed in a routine manner and enabled by causal (“hard wired”) mechanism that are independent of learning and control of the cognizer. However, learning from explicit rules leads to regress. Therefore, it seems that basic inferential steps are due to a blind “hard wired” mechanism, and successful inference depends on cognitive mechanisms that execute basic steps blindly and without much of conscious control by the cognizer. So the success of inferences is open to blind luck. The conceptualists insist on the role of understanding in great part in order to block the Achilles and the Tortoise conclusions. It is the luminous understanding of logical constants that meshes with awareness of rules. A tortoise that does not feel compulsion to infer in accordance with Modus Ponens also lacks the concept of implication, so the paradox is blocked.

Of course, a determined blind-compulsion theoretician will insist that the feeling of compulsion and of obviousness is equally pre-determined by a blind mechanism. Alternatively, a skeptic might direct our attention to the phenomenology of armchair thought and claim the following. It is possible to coherently imagine states of irrational quasi-understanding, (quasi-)inferring and (quasi-)proof-following that are phenomenally indistinguishable from rational understanding, inferring and proof-following. (C. Wright calls such activity “maundering.”) Due to phenomenological indistinguishability, it is impossible to tell from the first-person perspective whether one is in such an irrational state. Indeed, there is a fair chance that the actual pathology of thought exhibits such phenomena: some scenarios of such occasional attacks is not only possible, but seems to be actual. The beginning of “Beautiful Mind” reports the mathematician Nash as saying that he gets his intuition about extraterrestrials “from the very same source” from which he gets his mathematical intuitions. Therefore, even if one is *de facto* in the rational state, one cannot know this by reflection alone. Therefore one cannot in general *know* that one is undergoing an episode of rational understanding, rational inferring and proof-following, in contrast to their irrational counterparts. This is the argument from demon and madness, anticipated by Descartes, sketched by some of his interpreters, like, for instance H. Frankfurt and B. Williams, and developed by C. Wright (1991), the Dreamers and Madmen Argument, as we

might call it, borrowing from the title of Frankfurt's book. Of course, the problems created by the argument are common to both apriorist and aposteriorist, as the book-editors reminded me, but the domain addressed by it is the one especially dear to the heart of conceptualist-apriorist. Interestingly Peacocke does not address the challenge of this argument. This brings us to the work of C. Wright who does address it and who has been developing an interesting line of reply in the most original and detailed fashion. Following him, we shall introduce the third conceptualist justifier. Even if obviousness and compellingness create problems, we can neutralize them by appealing to this third source, he claims. We now turn to this claim.

3 Indispensability

Wright's answer to the problems has two components. First, he shows that the Dreamers and Madmen Argument in a way destroys or "implodes" itself. If the reasoner in the scenario is "maundering," she cannot be correctly inferring from her maunderings. So, the argument does not show that the thinker has no warrant.

If, as I earlier suggested, an effective argument from Dreaming, or from Brain-in-a-Vathood, etc., cannot proceed without all these elements—if our analysis does indeed capture the essential implicit detail of this kind of skeptical train of thought—then we may indeed draw a large but negative conclusion: *that there is actually no method of skeptically undermining our right to rely on any of our cognitive faculties using a fantasy, whatever its exact nature, of first-personally undetectable impairment.* (Wright 1991, 116)

We thereby conclude that ... it is not true that x has no warrant at t to believe that she is not then dreaming, and hence that the impossibility of earning a warrant to believe that one is not now dreaming—if that is what ... argument showed—does not imply that no such warrant is ever *possessed*. (107–108)

However, the Argument does suggest that the warrant cannot be earned by any kind of reasoning. The solution is to accept the

possibility of an *unearned* warrant. Applied to reasoning, the unearned warrant hinges upon the idea that logic is presupposed in every cognitive activity. Wright revives the Wittgensteinian conception of hinges and enriches it with his own idea of “cornerstones”:

Call a proposition a cornerstone for a given region of thought just in case it would follow from a lack of warrant for it that one could not rationally claim warrant for *any* belief in the region. The best—most challenging, most interesting—skeptical paradoxes work in two steps: by (i) making a case that a certain proposition (or restricted type of proposition) that we characteristically accept is indeed such a cornerstone for a much wider class of beliefs, and then (ii) arguing that we have no warrant for it. (Wright 2004a, 167–168)

If a cognitive project is “rationally non-optional,” i.e. indispensable in rational enquiry and in deliberation, then we may rationally take for granted the original presuppositions of such a project without specific evidence in their favor. The absence of defeating information is sufficient. So, Wright proposes to combine two warrant providers: the first is a rule-circular justification, reminiscent of justification from constitutiveness that answers the first order question: How might the knowledge of the validity of basic logical laws be arrived at? The second-order problem about such knowledge is “that of explaining with what right we claim it?” (Wright 2004b, 174) This “second-order problem has a chance of being addressed by invoking the notion of entitlement of cognitive project” (Ibid.).

This is of course the briefest of sketches. However, it is already visible that we have to do with a pale, etiolated version of classical transcendental arguments. But pallor is not in itself disqualifying. More important are problems that have to do with apriority. Wright claims that our acceptance of hinges and cornerstones is warranted by cognitive projects they enable. But this seems to take us from Wittgenstein to Quine and Putnam. If one’s early encounter with philosophy of logic was through Quine’s *Philosophy of Logic*, one would probably be reminded of the famous passages on universal use of logic. For Wright the acceptance of logic is warranted by its being indispensable to each and every cognitive project, each and every “region of thought.” Quine,

for his part, famously speaks of “lack of special subject matter: logic favors no distinctive portion of the lexicon, and neither does it favor one subdomain of values of variables over another.” And then passes to “the ubiquity of the use of logic. It is a handmaiden of all the sciences, including mathematics ... We might say at the risk of marring the figure that it is their promiscuity, in this regard, that goes far to distinguish logic and mathematics from other sciences. Because of these last two traits of logic and mathematics—their relevance to all science and their partiality toward none—it is customary to draw an emphatic boundary separating them from the natural sciences” (Quine 1970, 98). I recommend to take Quine’s quote with a Putnamian pinch of salt, as supporting a realist reading. Then, we seem to be back to the indispensability argument: logic and mathematics are to be accepted because they are indispensable to our widest cognitive projects in virtue of their admirable “versatile ancillarity.” The general background suggested by ancillarity and indispensability seems to be a means-end framework: logic and its acceptance are warranted as *means* for an end, not as valid in themselves, in stark contrast to the usual apriorist claims about autonomy.

Now, Cartesian skepticism “challenges knowledge of any external domain, whether abstract or spatiotemporal,” as Novák and Simonyi friendly suggested to me. Indeed, in the debate with the absolute skeptic no-one has a good chance. But the present point is different: if, instead of wanting to refute the skeptic in his own terms, you are involved in a “modest anti-skeptical project” that just aims to “set our own minds at ease,” to use the formulation of the well known contrast offered by Pryor (2000, 517), then the question of the comparative value of various indispensability considerations becomes interesting. And here Quine-Putnamian ones seem to be in better shape than Wittgenstein-Wrightian ones, simply because of being less restrictive.

However, we are not finished. There is an important contrast: in Wright’s view the justification is fully antecedent to the project, and this seems to guarantee apriority to logic. But does it really do it? Let us approach the right answer in a series of steps. First, is indispensability for *any* kind of large cognitive project by itself warrant-bestowing? To see that it is not, consider the following piece of reasoning. Belief in extraterrestrials is needed in order to embark upon a mega-project of re-interpreting a huge mass of recorded emission from outer space

as their messages. Therefore, if one has the project, one is warranted to believe in extraterrestrials. If you don't find the reasoning convincing, this suggests that the acceptance of hinges and cornerstones is justified by the meaningfulness of cognitive projects they enable, and is sensitive to it. The acceptance of logic is warranted by its being indispensable to each and every cognitive project, and is thus sensitive at least to the meaningfulness of major cognitive projects, and to the totality, say, "total inquiry" into what the world is like. But it is not clear that meaningfulness is completely independent of empirical considerations. Think of projects that would have sounded meaningful to an educated and intelligent Greek, or an Indian from the time of Mahabharata, and that do not sound meaningful to us any more.

Next consider the following considerations, which could be called the Mr. Magoo argument. Mr. Magoo has a very defective cognitive apparatus. His inductive propensities are idiotic, to use politically incorrect vocabulary, his senses most often deceive him, and his "heuristics" are ridiculous. (He lives in a super-hospitable environment, but hardly manages to survive.) His idiotic inductive propensities and ridiculous "heuristics," plus his misplaced uncritical trust in his senses are indispensable for his ever forming any belief. Therefore, he is warranted in taking them as unquestioned and unquestionable starting points.

If you don't find the reasoning convincing, this suggests that the acceptance of hinges and cornerstones is justified by the quality of cognitive projects they enable, and is *sensitive to the chances of their success*. But these chances are revealed by trying. Therefore, our best access to our own warrant involves information about the success of the relevant cognitive project. The warrant for logic is thus sensitive at least to the chances of success of our "total inquiry," and our awareness of it depends on the information about the success. Meaningfulness is not independent of chances of success, in virtue of "ought implies can" principle. Would this make the final justification just pragmatic? Well, it might. Wright has already been accused of smuggling in pragmatic considerations,²⁶ and the debate here is not about the absolute value of indispensability arguments but about the comparative status of the apriorist as against the aposteriorist line.

²⁶ For instance, by Duncan Pritchard (2005, section 3 of chapter 9).

Now, success of our “total inquiry” is to a large extent an empirical matter. Therefore, our awareness of it depends in the large measure of empirical information. Such information is *a posteriori*, which to the rationalist looks like a threat. How serious is it? On the aposteriorist line, some assumptions may be pragmatically antecedent to a cognitive project, but they are, firstly, justified by the success of the project, and secondly, revisable in the light of some advanced stage of the project. So, is the epistemic status of logic *a priori*, and can logic be revised on *a posteriori* grounds? First question first. We have seen that there is a touch of aposteriority present in the considerations of meaningfulness, and much more in our coming to know about our warrant. Now, officially, Wright can be unconcerned about it. It is the *having* of the warrant that is *a priori*, not knowing *about* it. But things are not that clear. First, this view seems to come very close to the aposteriorist view just mentioned: both theorists, our apriorist and our aposteriorist, accept a kind of warrant antecedent to the project, only give different names to the antecedence: the first takes it as bestowing serious apriority, the second, as bestowing only a kind of “vanilla” apriority, to use Harman’s ironical idiom (2003, title of section 1). Further, there is a problem about full reflective justification that goes beyond entitlement-warrant. How does the cognizer arrive at justified beliefs about herself being warranted? Well, partly by relying on the relative success of her total project. And this reflective justification might therefore be seriously *a posteriori*, in a way that precludes purely *a priori* justification of logic.

Notice that Wright is in no position to deny the need for full reflective justification. Here is his suggestion about the role of intellectual integrity and good conscience, which point precisely in the direction of reflective thought:

Descartes’ project in the *Meditations* was one of harmonization of his beliefs with the requirements of rational conscience and its timeless appeal is testimony to the deep entrenchment of virtues of intellectual integrity in our cognitive lives. The *right to claim* knowledge, as challenged by scepticism, is something to be understood in terms of—and to be settled by—canons of intellectual integrity. The paradoxes of scepticism are paradoxes for the attempt at a systematic respect of those canons. They cannot be

addressed by a position, which allows that in the end thoroughgoing intellectual integrity is unobtainable, that all we can hope for is fortunate cognitive situation. When good conscience fails, there are still, indeed, other good-circumstantial-qualities which our beliefs may have. But what is wanted is good conscience for the claim that this possibility is realized on the grand scale we customarily assume. (Wright 2004a, 211)

So, for Wright, the right to claim knowledge is inseparable from having good conscience for the claim that there is a realized possibility of knowledge on the grand scale. And this claim is clearly reflective. Now, if reflection on warrant involves important *a posteriori* elements, having to do with success of our cognitive project, then our full reflective justification is *a posteriori*.

The second question is one of *a posteriori* revisability. How much is written into the warrant? Does it preclude revision? Consider the analogy with the beliefs into the Material world: Here is what we are antecedently entitled to in the case of beliefs about material world, according to Wright: There is a material world, broadly in keeping with the way in which sense-experience represents it (2004a, 187). Now, this material-world hinge, as we might call it, is very vague. And it imposes very, very weak constraints. For instance, our sense-experience represents the material world as being to a large extent composed of solid, dense matter. But, if we discovered that the ultimate stuff were just “atoms and void,” this would not unhinge our sense-experience; nor would a still more dramatic revision, namely the discovery that it is fields of forces, very much *unlike* ordinary matter, that ultimately make up “material world.” In other words, the material-world hinge is *minimal*, it offers a broad umbrella statement, qualified with the clause “broadly in keeping,” that allows for dramatic revisions in interpretation. Exactly what one would expect from an *antecedent* assumption, open to all sorts of modifications.

Is there any reason built into the nature of entitlement why the logic-hinge should be any different? If its only *raison d'être* is indispensability for the cognitive project, then it is in the same boat with the material-world hinge, if not even less secure. Logic is needed for every “region of thought,” so it should be adaptable to each one of them as well. Is there any reason to think at this stage that our initial logic, the

one we find natural and obvious, is so universally applicable without a least revision? Of course not. Antecedently to experience, there is no reason to think that the initial logic, and our initial inability to imagine counterexamples to it, will be *that* successful. Our final rational confidence in logic might derive partly from the fact that it has never let us down, and this would be in keeping with its ancillary role, stressed by both Wright and Quine.²⁷

4 Summary and conclusion

We have discussed and criticized conceptualism, a dominant rationalist-apriorist line in recent debate about knowledge of logic and *a priori* justification, which relies upon human mastery of logical concepts which allegedly gives the cognizer access to “logical reality” and thus ultimately helps solving Benacerraf’s dilemma. We focused upon the versions offered by C. Peacocke and C. Wright. Conceptualists deny the need for or epistemological relevance of empirical explanation of our logical capacities, and offer three allegedly *a priori* sources of justification: first, the meaning- or concept-constitutive role of logical rules, second, their obviousness coupled with compellingness, and third, their indispensability. We have argued here that each of the proposed justifiers is beset with difficult problems. Let me re-state the problems and difficulties confronting the conceptualist account, and the three justification-warrant strategies its proponents employ. The common assumption is that most of our inferences concern worldly items, the semantic values of propositions employed and of their components, and that the beliefs about validity of inferences involve a certainty about such worldly items.

The first strategy, narrow justification through concept and meaning constitution confronts several problems. First, constitution problem: understanding logical constants might not be constitutive for having the concept that is its sense. Partial understanding might be sufficient, and understanding coupled with mistaken naïve assumption could be also sufficient for having the concept. A sophisticated

²⁷ I am leaving aside a related line that denies the need for justification or warrant, since it has not been prominent in accounting for apriority.

understanding coupled with exotic theories, either about logic or world or both, is certainly sufficient, and the exotic theories might lead the concept-owner to reject some links that are in fact constitutive of the concept in question.

Second, circularity problems: there is rule-circularity involved in justifying. There is an important domain-circularity in the attempt to justify following logical rules by intuitions concerning their semantics: rules and semantics are two aspects of largely the same subject matter, so the circle involved is very tight. On the other hand, conceptualists cannot appeal to a wider reflective equilibrium appealing also to the general success of logically correct inferences.

Third, the problem of determination. The cornerstone of conceptualist strategy is the idea that sense determines reference. But its application leads to a variant of Benacerraf's dilemma. If both sense and referent are fully objective, mind-independent items, there is less of a problem about determination: an objectively correct sense might be said to be coupled with objectively correct pattern of semantic values, and the issue of what determines what then becomes less urgent. But then the problem of grasping such Platonic, objectively correct senses reappears. How do such senses ever reach us? What explains our access to them? If, on the other hand, senses are more tied to the domain of the mental, if they are intrinsically mind-friendly, so to speak, then the opposite problems emerge. If such mind-bound senses metaphysically determine the pattern of semantic values, then logic-in-the-world is mind-dependent, and we end up in anti-realism or some non-realist variant of quietism. If the determination is only epistemic, i.e. if they merely guide us to the independent logical reality, the account is back to the square one: such guidance is precisely the thing to be ascertained and explained in the first place.

There is no wonder that conceptualists implicitly or explicitly agree that constitution is not sufficient for justification. This brings us to cognitive-epistemic justifiers and warrant providers. Consider first the appeal to obviousness and primitive compellingness coupled with anti-explanationism: constitution, obviousness and compelling nature of logical insight both account for knowledge of logic and justify logical beliefs and inferential practices. The classical problem concerning logical compulsion alone is the Achilles and the Tortoise problem: basic logical compulsion seems to be blind and thus a matter of sheer luck.

Since our reflection depends upon the very mechanism employed in unreflective reasoning, the luck cannot be avoided at the level of reflection. Appealing to obviousness of correct inferences might mitigate the predicament.

However, the classical problem for obviousness-cum-compellingness is the dilemma between a psychological and a normative-epistemological interpretation. If it is psychological, it leaves open the issue of objective validity. If it is purely normative, without inbuilt psychological ties, the opposite, Benacerrafian question arises: How does the thinker recognize the objective normative property in question?

A skeptical problem targeting the same justifier is the Dreamers and Madmen argument: Consider states of irrational quasi-understanding and inferring, phenomenally indistinguishable from their rational counterparts. It is impossible to tell from the first-person perspective whether one is in such irrational state. Therefore, even if one is *de facto* in the rational state, one cannot know this by reflection alone.

Next, a problem with the anti-explanationism. It seems to isolate the conceptual domain from the rest, placing some sort of veil of conception between the cognizer and the world she is thinking of. Concepts are indispensable, but their use should not be separated from issues of their origin and the source of their capacity to trace the truth about their domain. Traditionally, the issue of explanation has always been seen as part and parcel of the issue of reflective justification, the friendly explanations enhancing justification, and the unfriendly, for instance, the unmasking ones, taking it away. I see no reason to depart from the tradition.

Finally, there is the problem with the *a priori* status allegedly granted to logic by its indispensability for our cognitive projects, that makes logical beliefs into a cornerstone of our cognitive projects. The warrant inherited from projects is hostage to meaningfulness, and meaningfulness in turn partly depends on the project's potential for success, among other things. And all these warrant-providers are at least partly dependent on *a posteriori* considerations. Therefore, there is a significant *a posteriori* element to indispensability.

For a conceptualist philosopher in an optimistic mood this short budget of problems might look merely as a welcome challenge: Why doubt the capacity of conceptualism to find a satisfactory solution? But from a less optimistic standpoint, the aim of presenting them together,

albeit in an extremely sketchy form, is to point out that the program is beset with difficulties, and that simply adding new allegedly *a priori* justifiers to the list does little good. Given all these problems an alternative might be preferable, that would give more weight to *a posteriori* justifiers, and perhaps also to causal explanatory considerations, that might figure in a complete and reflective justification of our inferential practices and logical intuitions.

References

- Boghossian, P. 2000. "Knowledge of Logic." In Boghossian and Peacocke 2000.
 ———. 2003. "Blind Reasoning." *Proceedings of the Aristotelian Society, Supplementary Vol. 77*: 225–247.
- Boghossian, P. and C. Peacocke, eds. 2000. *New Essays on the A Priori*. Oxford: Clarendon Press.
- BonJour, L. 1998. *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
- Burge, T. 2003. "Reply to Peacocke." In *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, edited by M. Hahn and B. Ramberg. Cambridge, Mass.: MIT Press.
- Casullo, A. 2003. *A Priori Justification*. New York: Oxford University Press.
- Colyvan, M. 2008. "Indispensability Arguments in the Philosophy of Mathematics." In *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), edited by E.N. Zalta. <http://plato.stanford.edu/archives/fall2008/entries/mathphil-indis/> (accessed February 9, 2011).
- Dummett, M. 1991. *The Logical Basis of Metaphysics*. Cambridge, Mass.: Harvard University Press.
- Evnine, S. 2001. "The Universality of Logic: On the Connection between Rationality and Logical Ability." *Mind* 110 (438): 335–337.
- Harman, G. 2003. "The Future of the *A Priori*." In *Philosophy in America at the Turn of the Century*, APA Centenia Suppl. to *Journal of Philosophical Research*. Charlottesville: Philosophy Documentation Center.
- Hale, B. 1987. *Abstract Objects*. Oxford: Blackwell.
- . 2002. "Basic Logical Knowledge." In *Logic, Thought and Language*, Royal Institute of Philosophy Supplement 51, edited by A. O'Hear. Cambridge: Cambridge University Press.
- Hale, B. and C. Wright. 2000. "Implicit Definition and the *A Priori*." In Boghossian and Peacocke 2000.
- . 2002. "Benacerraf's Dilemma Revisited." *European Journal of Philosophy*, 10 (1): 101–129.
- Jeshion, R. 2001. "Frege's Notion of Self-Evidence." *Mind* 110 (440): 937–976.

- Miščević, N. 2000. "Quining the *A Priori*." In *Knowledge, Language and Logic: Questions for Quine*, edited by A. Orenstein and P. Kotatko. Dordrecht: Kluwer.
- Peacocke, C. 1987. "Understanding Logical Constants." *Proceedings of the British Academy* 73: 153–200.
- . 1992. "Sense and Justification." *Mind* 101(404): 793–816.
- . 1998. "Implicit Conceptions, Understanding and Rationality." *Philosophical Issues* 9: 43–88.
- . 1999. *Being Known*. New York: Oxford University Press.
- . 2000. "Explaining the *A Priori*: The Programme of Moderate Rationalism." In Boghossian and Peacocke 2000.
- . 2002. "Three Principles of Rationalism." *European Journal of Philosophy* 10 (3): 375–397.
- . 2004. *The Realm of Reason*. New York: Oxford University Press.
- . 2005. "The *A Priori*." In *The Oxford Handbook of Contemporary Philosophy*, edited by F. Jackson, and M. Smith. New York: Oxford University Press.
- Quine, W.V.O. 1970. *Philosophy of Logic*. Cambridge, Mass.: Harvard University Press.
- Pritchard, D. 2005. *Epistemic Luck*. New York: Oxford University Press.
- Pryor, J. 2000. "The Skeptic and the Dogmatist." *Noûs* 34 (4): 517–549.
- Rey, G. 1996. "Resisting Primitive Compulsion." *Philosophy and Phenomenological Research* 56 (2): 419–424.
- Ripps, L.J. 1994. *The Psychology of Proof*. Cambridge, Mass.: MIT Press.
- Russell, B. 1921. *The Analysis of Mind*. London: Allen and Unwin.
- . 1927. *An Outline of Philosophy*. London: Allen and Unwin.
- . 1940. *An Inquiry into Meaning and Truth*. London: Allen and Unwin.
- . 1948. *Human Knowledge: Its Scope and Limits*. London: Allen and Unwin.
- Sosa, E. 2003. "Ontology, Understanding and the *A Priori*." *Ratio* 16: 178–188.
- . 2004. "Two False Dichotomies: Foundationalism/Coherentism and Internalism/Externalism." In *Pyrrhonian Skepticism*, edited by W. Sinnott-Armstrong. New York: Oxford University Press.
- Stroud, B. 1979. "Inference, Belief, and Understanding." *Mind* 88 (350): 179–196.
- Williamson, T. 2006. "Conceptual Truth." *Proceedings of the Aristotelian Society, Supplementary Vol.* 80 (1): 1–41.
- Wright, C. 1991. "Skepticism and Dreaming: Imploding the Demon." *Mind* 100: 87–116.
- . 2004a. "Warrant for Nothing (and Foundations for Free)?" *Proceeding of the Aristotelian Society, Supplementary vol.* 78: 167–212.
- . 2004b. "Intuition, Entitlement and the Epistemology of Logical Laws." *Dialectica* 58 (1): 155–175.

IAN RUMFITT

What Is Logic?

1 Notions of consequence

What is logic? Textbooks typically introduce the subject as the science of consequence. Thus in an early section of his estimable primer—a section entitled “What Logic Is About”—we find Benson Mates explaining that

logic investigates the relation of *consequence* that holds between the premises and the conclusion of a sound argument. An argument is said to be *sound* (correct, valid) if its conclusion *follows from* or is a *consequence of* its premises; otherwise it is *unsound*.¹ (Mates 1965, 2)

In a similar spirit, E.J. Lemmon begins *Beginning Logic* by writing that

Only §5 of this essay overlaps substantially with my lecture in Budapest, the rest of which now strikes me as much ado about little. But the nature of logic is a topic of perennial interest, so I hope that the present piece will contribute something to the proceedings of an unusually pleasant and stimulating conference. The essay draws on material presented to seminars in Oxford and London, and on the first of my Nelson Lectures, given at the University of Michigan at Ann Arbor in September 2004. I am grateful to Jonathan Barnes, Dorothy Edgington, Crispin Wright, and the editors of this volume for their comments on a draft.

¹ Some logicians call an argument “valid” when its conclusion follows from its premises, and reserve the term ‘sound’ for valid arguments whose premises are true. I shall not follow them. The term ‘valid’ is tied in many people’s minds to the idea that there is a canonical, “logical” standard for validity; and this idea is one I wish to challenge.

logic's main concern is with the soundness and unsoundness of arguments ... Typically, an argument consists of certain statements or propositions, called its *premises*, from which a certain other statement or proposition, called its *conclusion*, is claimed to *follow*. We mark, in English, the claim that the conclusion follows from the premises by using such words as 'so' and 'therefore' between premises and conclusion ... Logicians are concerned with whether a conclusion does or does not follow from the given premises. If it does, then the argument in question is said to be *sound*; otherwise *unsound*. (Lemmon 1965, 1)

Both of these passages suggest that we have some pre-theoretical grasp of the relationship of one thing's following from some others. The laws of logic are then taken to say, in general terms, which things stand in this relationship. Thus a logical law which classical logicians accept, but which intuitionist logicians do not accept without restriction, says that a proposition follows from the negation of its negation. Many other passages could be cited which express this view of logic's business.

There is, to be sure, one important matter over which adherents of the view differ: the nature of the *relata* of the consequence relation. Lemmon writes of one statement's following from some others; by a statement he means something which is stated, or which could be stated, by the utterance of a declarative sentence on a given occasion of use (1965, 6). This way of speaking is undeniably natural: it comes easily to say "The statement that every set can be well ordered follows from the statement that there is a choice function on every set." There are, though, problems with this account of the *relata* of consequence, at least if it is deployed early in an investigation into the nature of logic. If consequence is a relation among objects, then its *relata* must be subject to the discipline of the identity relation, for such subjection is the mark of objects. There are reasons to doubt, though, if our ordinary standards for assessing whether two utterances "say the same thing" can really sustain judgments of strict identity between what Lemmon calls statements.² A theorist may of course try to supplement those ordinary standards with additional criteria that can sustain such judgments, and I have no argument to show that this attempt must fail.

² I set out these reasons in my essay "Objects of Thought" (Rumfitt forthcoming).

It seems certain, however, that any attempt of this kind will invoke a number of logical laws, and so is best delayed (if possible) until after we have made some progress in elucidating the nature of those laws.

And delay is possible if we proceed as follows. Let us consider those ordered pairs whose first element is a meaningful, indeed disambiguated, declarative type sentence, and whose second element is a possible context of utterance; by a possible context of utterance, I mean a determination of all the contextual features which can bear upon the truth or falsity of a declarative utterance. Some of these ordered pairs will be such that, were the declarative type sentence that is the first element uttered in the context that is the second element, a complete thought would then be expressed. As I shall use the term, a *proposition* is an ordered pair that meets this condition, and I take propositions (in this sense) to be the *relata* of consequence relations. Not every ordered pair of declarative type sentence and possible context of utterance will qualify as a proposition. For example, an ordered pair of sentence and context whose first member is 'You are ill' will not count as a proposition unless the context supplies an addressee. On this way of understanding the term, each proposition belongs to a language, namely, the language of the sentence that is its first element; and the English proposition "Every set can be well ordered" is clearly distinct from the proposition "There is a choice function on every set": the propositions have distinct type sentences as their first members. Furthermore, each proposition possesses a sense or content: this will be the thought that would be expressed by uttering the proposition's first element (the declarative sentence) in the context that comprises its second element. It then makes sense to classify a proposition as true or false *simpliciter*, according as that thought is true or false. Both assertions and utterances made within the scope of suppositions may be instances of propositions in the present sense, so our theory allows us to consider logical relations among suppositions. As we shall see, this is important. It also explains why I prefer the term 'proposition' to Mates's and Lemmon's 'statement': while one may propound without asserting, it would be infelicitous to allow unasserted statements, for 'states' often means 'asserts.' It should be remarked, though, that my use of the term 'proposition' differs from the one that prevails in contemporary philosophy: I do not take a proposition to be what a declarative utterance expresses.

Despite their differences over the nature of premises and conclusions, Mates and Lemmon are at one in appealing to an antecedent understanding of the relation of consequence in characterizing the logician's task. When Mates describes logic as investigating *the* relation of consequence that holds between the premises and the conclusion of a sound argument, he implies, or presupposes, that there is some single, uniquely favored relation of consequence which has a special claim on the logician's attention. Similarly, Lemmon presupposes that we shall know what he means by 'following' when he says that "logicians are concerned with whether a conclusion does or does not follow from the given premisses." But do we know this? The glosses philosophers and logicians have placed on 'follows from' are of limited help in identifying the intended relation. In the passage in which he appropriated the word 'entails' from the lawyers, G.E. Moore laid it down that we shall "be able to say truly that ' p entails q ' when and only when we are able to say truly that ' q follows from p ' ... in the sense in which the conclusion of a syllogism in Barbara follows from the two premisses, taken as one conjunctive proposition; or in which the proposition 'This is coloured' follows from 'This is red'" (Moore 1922, 291). Even after studying logic, however, one may be forgiven for doubting that there is a single such sense, or a single relation of properly logical entailment. For if there is, then no logical system that I know of comes close to characterizing it completely.

In §6 below, I will identify a relation that deserves the title of "logical" consequence in the "broad" sense that Moore seems to have in mind. I deny, though, that we have a firm grasp on such a notion in advance of theorizing about logic, so it is a mistake to describe the logician as setting out to investigate one such relationship. I shall begin by justifying my claim that this is a mistake and showing how its diagnosis leads to a better account of logic's business, an account which eventually leads to a principled identification of broadly logical consequence.

Mates and Lemmon both postulate a connection between an argument's being sound and its conclusion's following from its premises. Some such connection surely obtains, but we should recognize that our ordinary standards for assessing the soundness of arguments—and hence our ordinary standards for saying what follows from what—vary from context to context. A physicist who argues "This body is accelerating; so a force must be acting on it" may be deemed to have argued

soundly; and his conclusion may be said to follow from his premise. But that is because the relationship of consequence that sets the standard for assessing the soundness of arguments within physics excludes physical impossibilities as irrelevant to consequence. A lawyer who argues “Jones gave Smith £10 in return for his undertaking to deliver certain goods by 1 October; Smith failed to deliver those goods by that date; so Smith is liable to compensate Jones for the losses he incurred because of that failure” may also be said to have argued soundly; and his conclusion may be said to follow from his premises. But that is because the salient relationship of consequence in this context excludes as irrelevant to consequence anything excluded by the laws and the precedents of the pertinent jurisdiction. Indeed, things far more ephemeral than the principles of physics, or the laws and legal precedents of a given state, may generate consequence relations that underpin assessments of soundness. As Timothy Smiley has observed, a traveler who reasons “It’s Tuesday; so it must be Paris” may also be said to have argued soundly (Smiley 1995, 725–736). This is because the cognate consequence relation excludes as irrelevant possibilities that conflict with the traveler’s timetable.

Those who suppose that we have some pre-theoretical grasp of a favored relationship of logical consequence recognize that our standards for assessing the soundness of ordinary arguments vary from case to case. But they typically try to explain away our tendency to classify as sound arguments that do not meet their canonical standards of “logical validity” by invoking the notion of a tacit or suppressed premise. Thus Copi:

Because it is incomplete, an enthymeme must have its suppressed premise or premises taken into account when the question of its validity arises. Where a necessary premise is missing, the argument is technically invalid. But where the unexpressed premise is easily supplied and obviously true, in all fairness it ought to be included as part of the argument in any appraisal of it. In such a case one assumes that the maker of the argument did have more “in mind” than he stated explicitly. In most cases there is no difficulty in supplying the tacit premise that the speaker intended but did not express. Thus “Al is older than Bill. Bill is older than Charlie. Therefore Al is older than Charlie” ought to be counted

as valid, since it becomes so when the trivially true proposition that *being older than* is a transitive relation, is added as an auxiliary premise. (Copi 1973, 132)

It is sometimes right to appraise an argument as though it contained an “additional” premise—a premise which its proponent omitted to articulate. But the strategy of postulating unexpressed premises does not provide a satisfactory general explanation of our intuitive assessments of the soundness of arguments. Those assessments rest in part on our ability to latch onto the consequence relationship that is relevant in the context of argument. So the strategy will only work if any consequence relationship we can readily latch onto factors into a relationship of properly logical consequence, along with the hypothesis that such-and-such propositions are available to serve as suppressed, or unexpressed, premises.

Is this condition met, though? If we follow Copi in supposing that the favored consequence relation is that of classical first-order logic, then it is not. Consider the usual proof within real analysis of Archimedes’ law: given any two positive real numbers p and q , there is a natural number m such that the product of m and p is greater than q ; there are, in other words, no “infinitesimal” real numbers. The proof of the law is simple. One considers the minimal closure of the singleton $\{p\}$ under addition; then take C to be the property of a being a real number belonging to this closure. It is then easy to show that C cannot have a least upper bound. The completeness axiom for the real numbers says that whenever a property of real numbers has an upper bound, it has a least upper bound. So by contraposition C has no upper bound, and the law follows.

Simple as the proof is, however, the consequence relation that we latch onto when we follow it cannot be represented as first-order consequence given a set of suppressed additional premises. It cannot be represented in this way even if the set of additional premises is allowed to be infinite.³ The problem lies in the way the completeness axiom quantifies over all the properties of real numbers. In a first-order logic,

³ Smiley has objected to the enthymematic strategy that there may be no way of capturing the generality of an extra-logical rule in a single proposition (1995, 732). The observation is correct, but it invites a revised strategy in which the “tacit” premises may instead be supplied as instances of an axiom schema. The present objection tells against the revised strategy.

that quantification will have to be formulated schematically. The axiom will be understood to comprise all the instances of the schema ‘when-ever there is an upper bound for P , there is a least upper bound for P ,’ where the schematic letter ‘ P ’ is replaced by a one-place predicate. But many properties of real numbers (even simple minimal closures such as C) are expressed by no predicate, so the proof does not go through if the completeness axiom is construed as a schema. And there are indeed non-Archimedean models of first-order analysis.

It may be replied that this objection shows only that Copi chose the wrong logic with which to execute his enthymematic strategy. The objection would lapse if the background logic were full second-order logic. But that reply raises another problem. No doubt we can latch onto the second-order consequence relation, but we can also latch onto the relation of first-order consequence. And the latter relation cannot be characterized as second-order consequence, given a set of suppressed additional premises. To obtain first-order from second-order consequence, one needs to impose a syntactic restriction, forbidding quantification into the places occupied by predicates—a very different matter from postulating premises.

There is a second, more philosophical, objection to the enthymematic strategy: the lack of any persuasive motivation for butchering the surface structure of arguments in the way that it requires. As I have said, our ordinary assessments of argumentative soundness rely on our ability to latch onto the consequence relationship that is relevant in the argument’s context. Having latched onto that, we can appraise many arguments more or less as they come. No doubt some are best appraised by postulating an additional premise which the proponent intended but did not express. Equally, though, there will be many contexts in which we do not need to postulate any additional premise in order to account for the soundness of “Al is older than Bill. Bill is older than Charlie. Therefore Al is older than Charlie.” We need to ask, then, what philosophical principle is supposed to sustain the claim that this argument is “strictly speaking” unsound as stated, and needs to be supplemented by some premise expressing the transitivity of *being older than*.

The features of logical consequence that are usually cited to mark out its special status do not by themselves sustain this claim. Some philosophers like to say that the logical consequences of premises are “implicit” in them. What they mean is murky, but on any reasonable

explication, Al's being older than Charlie is implicit in his being older than Bill and Bill's being older than Charlie. (Certainly, these premises necessitate Al's being older than Charlie in any natural sense in which premises of an application of *modus ponens* necessitate its conclusion.) Others will say that failure to accept at least the obvious logical consequences of what one says is a sign that one does not properly understand it. But the same goes for any of a proposition's obvious consequences, whether or not those consequences are deemed to follow "logically."⁴ In particular, someone who accepts that Al is older than Bill, and that Bill is older than Charlie, while failing to accept that Al is older than Charlie gives a sign that he does not understand comparative adjectives. As for the idea that "logical validity is special because, being independent of external circumstances, it can be assessed simply by looking at the words and sentences involved and the way they relate to one another,"⁵ it fails to draw the distinction in the intended place: someone can assess our argument about Al, Bill and Charlie simply by looking at the words, and without discovering the men's ages. Perhaps, indeed, the idea fails to draw a distinction at all. The postulated distinction between the conceptual and the empirical was a target of Quine's assault on the dogmas of empiricism (Quine 1951). A thinker with a sufficient grasp of the words 'force' and 'acceleration' may be able to recognize the soundness of "This body is accelerating; so a force is acting on it." But that hardly shows that the soundness of this argument is "independent of the exter-

⁴ A cognate point is made by Quine, as he argues for the "emptiness" of the positivist doctrine that a logically true proposition is true by virtue of its meaning alone. For example: "there can be no stronger evidence of a change in usage than the repudiation of what had been obvious, and no stronger evidence of bad translation than that it translates earnest affirmations into obvious falsehoods" (1976, 113). Despite his occasional descriptions of logic as "the science of necessary inference," I read Quine as a fellow skeptic about the claim that we have a clear pre-theoretical notion of logical consequence: "There are philosophers of ordinary language who have grown so inured to the philosophical terms 'entails' and 'inconsistent' as to look upon them, perhaps, as ordinary language. But the reader without such benefits of use and custom is apt to feel, even after Mr Strawson's painstaking discussions of the notions of inconsistency and entailment, somewhat the kind of insecurity over these notions that many engineers must have felt, when callow, over derivatives and differentials. At the risk of seeming unteachable, I go on record as one such reader" (Quine 1976, 138).

⁵ Smiley 1995, 733. Smiley does not himself accept the idea.

nal circumstances.” On the contrary: it depends on whether the external circumstances conform to Newton’s laws of motion.

2 Logic as comprising the general laws of consequence

These considerations make it doubtful whether we have a secure pre-theoretical grasp of a favored relation of logical consequence whose more precise investigation is the logician’s business. Perhaps reflection on logic will enable us to identify such a relation, but this identification cannot come at the beginning of an account of what the logician is about. As I now argue, though, this conclusion engenders no mystery about his business. It merely shows that a proper description of that business must acknowledge the plethora of relations of consequence.

In order to find such a description, we may begin by remarking that certain arguments stand or fall together. If the argument (*A*)

This body is accelerating
So: a force is acting on it

is sound (by certain contextually determined standards) then the argument (*B*)

No force is acting on this body
So: it is not accelerating

is also sound (by those very same standards). This is because if the conclusion of (*A*) stands in a certain consequence relation to the premise of (*A*), then the conclusion of (*B*) stands in the same consequence relation to the premise of (*B*). What is more, arguments exhibiting the pattern of relationship as (*A*) and (*B*) also stand or fall together, even when they are to be appraised by a very different relationship of consequence. If the argument (*C*)

It is Tuesday
So: this is Paris

is sound (by given standards), then so will be the argument (*D*):

This is not Paris
 So: it is not Tuesday

by those same standards. The consequence relation that underpins the soundness of (C) will be very different from that which underpins the soundness of (A) and (B) . For all that, we can still say that if the conclusion of (C) stands in that very different relation to the premise of (C) , then the conclusion of (D) will also stand in that relation to the premise of (D) .

These examples point towards a conception of logic on which its laws do not characterize a single relation of logical consequence. Rather, they suggest that we can and should take those laws to be general truths concerning all relationships of consequence. Thus we may say that there is a logical law which accounts for the facts that the soundness of (A) (in a given argumentative context) stands or falls with the soundness of (B) (in that same context), and that the soundness of (C) similarly stands or falls with the soundness of (D) , and so forth. A first shot at formulating that law might be:

- (1) Whatever relationship of consequence R might be, if a proposition B stands in R to a proposition A , then any negation of A stands in R to any negation of B .

Formula (1) expresses, in the present framework, what is usually called the law of contraposition. The law is untypically simple, for, in general, one term of a consequence relation will be plural: a conclusion will follow, or fail to follow, from one or more premises. A consequence relation obtains *among* propositions, as I shall say, not simply between them.

Some may be surprised by one feature of the formulation of contraposition. Our formula speaks of any negation of a proposition, whereas most logicians hold, or presuppose, that each proposition has one and only one negation. On our account of what a proposition is, however, such a supposition is untenable. “Not every man is mortal” and “It is not the case that every man is mortal” are distinct English propositions but each is a negation of the proposition “Every man is mortal.” Of course, when a logician speaks in the singular of *the* negation of a proposition, he does not mean to deny this, but quite what he

means to assert is a delicate matter. Our formulation, which makes no presumption of uniqueness, brackets the issue.

Logical laws such as (1) may seem *recherché*, but they can be brought into a familiar form. If we use the colon to signify an arbitrary relation of consequence, and use ' $\neg A$ ' to signify any negation of A , then law (1) may be formalized as

$$\text{If } A : B \text{ then } \neg B : \neg A.$$

Where a long horizontal line signifies “infer,” this law corresponds to the following rule of proof:

$$\frac{A : B}{\neg B : \neg A.}$$

This rule is a familiar rule of a sequent calculus. More exactly, it is a rule of a restricted sequent calculus in which one and only formula may appear on the right of the colon (i.e. in which the “succedent” is restricted to a single formula).⁶ I do not claim that Gentzen had this interpretation in mind when he showed how to formalize classical and intuitionistic logic as systems of sequent calculi (Gentzen 1935). So far as I can see, though, nothing in that formalization requires that the colon should be taken to signify some one, fixed notion of logical consequence. If, instead, we read it as signifying an arbitrary relation of consequence, then the rules of the sequent calculus correspond to what I am calling logical laws. Those rules are more general than many familiar laws. Thus, in a sequent calculus in which one and only one formula appears on the right of the colon, the standard rule for introducing the conditional on the left of the colon is:

$$\frac{X, B : C \quad Y : A}{X, Y, A \rightarrow B : C}$$

where each of A , B and C is an arbitrary formula and where X and Y are arbitrary sets of formulae. In the special case where X is empty, and

⁶ I justify this restriction in §5 below.

Y is a singleton whose only member is A , and where C is identical with B , this reduces to

$$\frac{B : B \quad A : A}{A, A \rightarrow B : B}$$

If we assume that for every consequence relation, a proposition will stand in that relation to itself, the conditions above the line will be fulfilled no matter what consequence relation the colon may signify. The special case, then, corresponds to the law of *modus ponens*:

- (2) Whatever relationship of consequence R might be, a proposition B stands in R to any pair of propositions comprising A together with the conditional proposition whose antecedent is A and whose consequent is B .

The additional generality, though, does not gainsay my claim: a familiar way of formalizing logic may be understood to embody the recommended conception of logical laws.

3 What is a consequence relation?

This is fine as far as it goes, but it makes one question pressing. On the view I am recommending, the laws of logic do not characterize some single relation of logical consequence. Rather, they are general laws of relations of consequence. But what marks out consequence relations (as I shall henceforth call them) from other relations among propositions?

The basic mark of a consequence relation is this: whenever it obtains between some premises and a conclusion, the truth of the premises *guarantees* the truth of the conclusion. This principle respects the connection between consequence and argumentative soundness. When we deem an argument sound, we are not saying merely that the truth of the premises makes the truth of the conclusion very likely. Nor even are we saying that the truth of the premises provides the strongest evidence that could be offered for the truth of the conclusion, as when we

make a prediction on the strongest possible inductive grounds. Rather, we are denying that there is the slightest possibility of the premises being true and the conclusion's not being true.

It is natural to gloss 'guarantee' using the modal notion of a possibility, but we need to recognize that the space of relevant possibilities varies from context to context. This variation is what generates the plethora of consequence relations. Consider again our physicist's argument: "This body is accelerating; so a force is acting on it." The conclusion follows from the premise because, in the context of the argument, the space of relevant possibilities is confined to physical possibilities; and there is no physical possibility in which the premise is true and the conclusion untrue. Perhaps other kinds of possibility would allow the premise to be true and the conclusion not to be true. But these possibilities do not threaten the claim of consequence, for the argument's context excludes them as irrelevant. Similarly, we may deem the package tourist's syllogism sound in an argumentative context in which the only relevant possibilities are those that conform to his timetable.

One may indeed say that to each space of possibilities, S , there corresponds a consequence relation R as follows:

(*Cons*) Some premises $A_1, \dots, A_n$ R -relate to a conclusion B if and only if, for any possibility x in S , if $A_1, \dots, A_n$ are all true in x then B is true in x too.

A proposition is deemed to be true *in* a possibility when it would be true, were the possibility to obtain. As we saw, assessing an argument's soundness involves latching on to the consequence relation that determines soundness. By (*Cons*), latching onto that consequence relation is in turn a matter of identifying the relevant space of possibilities.

A few remarks on (*Cons*) and the notions involved in it are in order.

By a possibility, I mean a way things might be, or might have been. But I mean *just* that: not a fully determinate way that all things might be, or might have been. So 'possibility' in (*Cons*) does not mean 'possible world,' even in Stalnaker's sense. A possible world is required to be complete: for any possible world w and any proposition A , A must either be true in w or false in w (deeming a proposition to be

false when its negation is true). Nothing comparable holds for possibilities. There is a possibility in which a child has just run from the house; things—some things anyway—might have been that way. But the proposition that a boy just ran from the house is not true in this possibility: we are not entitled to assert that the proposition would be true, were the possibility to obtain, i.e. were things that way. But equally, the proposition is not false, if being false amounts to having a true negation. For neither are we entitled to say that the proposition's negation would be true, were the possibility to obtain. So there are possibilities in which certain propositions are neither true nor false. This thesis need involve no departure from classical logic or semantics: the disjunctive proposition "Either a boy ran from the house or it is not the case that a boy ran from the house" is true in the relevant possibility. The point is simply that some possibilities are incomplete in that they do not determine the truth or falsity of certain propositions.

Second, (*Cons*) invokes the notion of a proposition's being true in a possibility. In applying this notion, we are to take the proposition in its actual sense, and ask whether it would be true were the relevant possibility to obtain.⁷ Thus (*Cons*) yields notions of consequence that are "representational" rather than "interpretational" in John Etchemendy's sense.⁸ Whether a conclusion follows from some premises depends on whether there is a relevant possibility in which they are true but it is not. In applying this test, we take the propositions' sense as given and consider their truth in all the relevant possibilities. We do not take the facts as given and consider all the different possible re-interpretations of the propositions' non-logical vocabulary. Etchemendy's book presents a sustained and persuasive argument for the superiority of representational accounts of consequence over the interpretational accounts, notably the famous theory of consequence due to Tarski (1936).

All the same, (*Cons*) vindicates some of Tarski's other claims about consequence. Before he essayed his account of logical consequence, Tarski in effect laid down three axioms which any consequence relation

⁷ I shall say more about the notion of a proposition's being true *in* a possibility in §7 below.

⁸ See Etchemendy 1990, chaps. 2 and 4.

must validate.⁹ Any such relation must be *reflexive*: a conclusion follows from any premises of which it is one. It must also be *monotonic*: if a conclusion follows from some premises, it also follows from any premises of which they are some. And any consequence relation must exhibit the form of transitivity that logicians call the “cut” law: if a conclusion follows from some premises, and if each of those premises in turn follows from some further premises, then the original conclusion follows from the further premises. Any relation satisfying (*Cons*) will satisfy each of these axioms. So, for example, if the proposition B is one of the propositions $A_1, \dots, A_n$, then for any space of possibilities S it will certainly hold good that for any possibility x in S , if $A_1, \dots, A_n$ are all true in x then B is also true in x . So for any relation R satisfying (*Cons*), we have that $A_1, \dots, A_n$ R -relate to B whenever B is one of $A_1, \dots, A_n$, showing that any such relation is reflexive.¹⁰ The arguments for monotonicity and transitivity are similar. Tarski never claimed that his axioms completely captured the ordinary notion of a consequence relation, and he would have been wrong to do so. The ordinary notion corresponds to ordinary evaluations of arguments as sound or not, and there will be many relations among propositions that respect Tarski’s axioms without yielding anything recognizable as a variety of argumentative soundness. But his axioms are still important in that they articulate features which mark out sound arguments from arguments that are good in other ways. So, for example, inductive support is not monotonic. Premises saying that a large number of pure samples of bismuth melt at 271°C support the generalization that all pure samples of bismuth melt at that temperature; not so if those premises are supplemented by others saying that further pure samples melt at 261°C. Similarly, the relation that obtains when some premises make a conclusion probable to a specified degree is not transitive. The truth of A might make that of B 99% probable, and the truth of B might make that of C 99% probable, without the truth of A making that of C 99% probable.

⁹ I say “in effect,” for his axioms characterize the functor ‘consequence of X ,’ where X is a set of declarative sentences, rather than any relation. For present purposes, though, that difference may be ignored. See Tarski 1930, 22–29, esp. 22–23.

¹⁰ This justifies the assumption used in §2 to derive law (2) (*modus ponens*) from the sequent rule for introducing $\rightarrow$ on the left.

4 Why consequence involves modality

Even though (*Cons*) validates these familiar and reassuring features of consequence relations, some philosophers will object to its explicating consequence using a modal notion (in this case, possibility). It is in place to consider the most famous such objection, that of Russell.

Russell argues as follows:

In order that it be *valid* to infer q from p , it is only necessary that p should be true and that the proposition “not- p or q ” should be true. Whenever this is the case, it is clear that q must [*sic*] be true. But inference will only in fact take place when the proposition “not- p or q ” is *known* otherwise than through knowledge of not- p or knowledge of q . Whenever p is false, “not- p or q ” is true, but is useless for inference, which requires that p should be true. Whenever q is already known to be true, “not- p or q ” is of course also known to be true, but is again useless for inference, since q is already known, and therefore does not need to be inferred. In fact, inference only arises when “not- p or q ” can be known without our knowing already which of the two alternatives it is that makes the disjunction true. Now, the circumstances under which this occurs are those in which certain relations of form exist between p and q ... But this formal relation is only required in order that we may be able to *know* that either the premise is false or the conclusion is true. It is the truth of “not- p or q ” that is required for the *validity* of the inference; what is required further is only required for the practical feasibility of the inference. (Russell 1919, 153)

In other words, *material* or *Philonian*¹¹ consequence—the relation that obtains when either the conclusion is (actually) true or some premise is (actually) untrue—suffices for what I have been calling an argument’s soundness. Russell proceeds to castigate C.I. Lewis for invoking modal

¹¹ “Philo [of Megara] says that a sound conditional is one that does not begin with a truth and end with a falsehood, e.g. when it is day and I am conversing, the statement ‘If it is day, I am conversing’” (Sextus, *Pyrhoneiae Hypotyposes*, ii. 110, as translated in Kneale and Kneale 1962, 128). ‘Sound conditionals’ translates the Greek dialecticians’ term for what we would now call a correct sequent.

notions in explicating consequence (154). If his argument were sound, it would tell equally against (*Cons*).

What makes Russell's argument interesting is that it contains an important insight. An inference, he tells us, requires that the premise should be true and that the conclusion should not be known. This is scarcely comprehensible unless we take 'inference,' as it comes from Russell's pen, to mean 'an instance of reasoning in which a thinker comes to know a conclusion by deducing it from premises that he knows already.' So the argument tacitly presents as the focal cases of logical appraisal those instances of argumentation in which a thinker purports to use his capacity for deductive reasoning to gain new knowledge from old.

I think that such cases should indeed be the *foci* of logical appraisal. Certainly, the value of our capacity for deductive argument lies in the fact that by exercising it we can gain knowledge that we could not otherwise obtain. Suppose I am strapped to the chair in my study. From that chair, I cannot see the street below. I do, however, see that it is raining, and thus know that it is raining. Moreover I know, ultimately on inductive grounds, that if it is raining the street is wet. Accordingly, I reason as follows:

1. It is raining
2. If it is raining, the street is wet
- So 3. The street is wet.

My reasoning here is a simple example of what Russell calls an inference. And the case brings out how, by exercising a deductive capacity, a thinker can gain knowledge—knowledge that he would not have been able to gain otherwise. In the case described, I know that it is raining by virtue of seeing that it is raining. And I know through induction that if it is raining, the street is wet. By making the deduction, I thereby come to know that the street is wet. *Ex hypothesi*, I cannot see that the street is wet, so I cannot come to know the conclusion by exercising my perceptual capacities, which is how I came to know the first premise. Similarly, I cannot come to know the conclusion on inductive grounds alone, which is how I came to know the second premise. Even in England, so pessimistic view of the weather (or of the wastefulness of the water companies) would not yield knowledge. But by exercising

my deductive capacity on the knowledge that is delivered by perception and induction, I can come to know something that I could not know on either of those bases by itself.¹²

How is it, though, that my conclusive belief—my belief that the street is wet—has the status of knowledge? Let us call a thinker *deductively capable* (with respect to a given consequence relation) if he deduces a conclusion from some premises only when the conclusion really does stand to the premises in the relevant consequence relation, and if he also recognizes at least some of the more obvious instances of that relation. And let us suppose that I am deductively capable with respect to whatever consequence relation the context makes relevant to the evaluation of our simple inference. Now because I know the two premises of that inference, those premises are true. And because the premises are true, and the conclusion follows from them, the conclusion is also true. *Ex hypothesi*, I am deductively capable, so I will deduce a conclusion from some premises only if it really does follow from them. Accordingly, when I deduce a conclusion from premises I know, the belief thereby formed will be true, and will have been produced in a way that reliably yields true beliefs. That belief, then, has the status of knowledge, or something close to it. This explanation, it may be noted, does not presume that I know *that* “The street is wet” follows from “It is raining” together with “If it is raining then the street is wet.” Indeed, it does not presume that I have the concept of consequence at all. Being able reliably to trace what follows from premises comes first, and is all we need to know things by deduction. Conceiving of propositions as standing in relations of consequence comes later, when we start to do logic, if it comes at all.

The argument of the previous paragraph does, however, rest on the factivity of “knows”: this is what underwrites the inference from the thinker’s knowing the premises to their being true. So it does not extend to explain why a deductive capacity would enable a thinker to extend warrants or justifications for premises that fall short of knowledge to comparable warrants or justifications for conclusions that are correctly deduced from those premises. Senses of ‘warrant’

¹² There remains the problem, which preoccupied J.S. Mill, of explaining *how* our capacity for deductive argument can yield new knowledge. I venture some remarks about this problem in my essay “Knowledge by Deduction” (Rumfitt 2008).

or ‘justification’ that fall short of knowledge will precisely not be factive, so the corresponding arguments for them would break down. But this, I think, is as it should be: knowledge is special in this regard. One might at first expect a deductive capacity to enable a thinker to splice together different sources of warrant or justification to produce a new warrant, even if the component warrants fall short of knowledge. But in general this latter kind of splicing is not possible. I may have warrant or justification for the claim that Charles is now in Paris: he always stays there at this time of year, I saw him heading off to the Eurostar, etc. I may also have warrant or justification for the claim that Charles is not in Paris: I have just received a postcard, apparently sent by him from Berlin. If I notice the conflict of evidence, I may make further investigations to resolve it. But what I will not do—what I cannot do—is to apply ‘and’-introduction to splice my various warrants or justifications together, so that collectively they support the claim that Charles both is and is not in Paris. For nothing at all warrants or justifies that. So far from combining to warrant the conjunctive conclusion, my justifications for the two premises undermine each other. It is a merit of our analysis that it brings out why a logical capacity need not enable one to extend warrants or justifications that fall short of knowledge, and thereby shows what is special about knowledge. The argument of the previous paragraph does presume, though, that the relevant possibilities do not change in the course of the reasoning, so we need not gainsay those who hold that such arguments as “This is a zebra; so this is not a mule cleverly disguised to look like a zebra” can take one from a known premise to an unknown conclusion (Dretske 1970). For their ground for holding this is precisely that the relevant space of possibilities expands between the premise and the conclusion. Whether the familiar skeptical tropes exploit such variation in the space of possibilities is a question in epistemology that we need not address here.

Russell was right, then, to focus on cases in which a thinker exercises his deductive capacity on premises he knows and thereby attains knowledge of a conclusion. And our account of logic then explains why the subject is useful. As we have seen, a thinker’s exercise of a deductive capacity enables him to extend his knowledge. So any thinker will benefit from mastering generally applicable techniques for extending his deductive capacities. On my conception of the subject, learning logic affords such techniques. In learning about physics, say, a

thinker may start with a rather limited deductive capacity: it could even be the minimal capacity that consists in the ability reliably to trace the consequence relation which is the closure under Tarski's axioms of the relation that obtains between the premise and conclusion of arguments isomorphic with (A). But if the thinker can reliably contrapose, then he will also come to command a wider relation that encompasses the premises and conclusions of arguments isomorphic with (B). So similarly for other logical rules. What is more, mastery of contraposition and the other rules will also expand his command of consequence relations in other fields—indeed, in *all* other fields, given that the logical particles such as 'not' and 'all' are ubiquitous. The theorems of logic may convey no substantive information, but they are simply the uninteresting by-products of rules, mastery of which expands a thinker's various deductive capacities—capacities whose exercise enables him to expand his substantive knowledge. Logical techniques are no less valuable for being indirectly useful, and the subject is no less valuable for being a second-order discipline.

All the same, Russell's conception of how knowledge is attained by applying a deductive capacity is too narrow, and when we widen that conception as we are in the end forced to widen it, we see that his strict Philonianism cannot be sustained. Deduction is useful, Russell tells us, in cases where we know that either not *P* or *Q* without knowing either disjunct. The question to press on him is *how* a thinker might know that either not *P* or *Q* without knowing that not *P* or knowing that *Q*.

Clearly, there are many ways in which a thinker might attain this combination of knowledge and ignorance. One way is through testimony: a trustworthy physicist may tell me that a given body is either not accelerating or has a force acting on it without telling me which. In order to explain how knowledge of this disjunction combines with knowledge that the body is accelerating to yield knowledge that a force is acting on it, we shall not need to suppose that the truth of the premises in any sense necessitates that of the conclusion.

For this reason, I think we should grant to Russell that the Philonian relation is a respectable consequence relation. But it is far from being the only such relation. We can come to know that either not *P* or *Q* without knowing that not *P* or knowing that *Q* otherwise than by testimony, and I claim that any full explanation of this combination of knowledge and ignorance will at some point advert to our ability to apply our

deductive capacities to propositions that are entertained as *suppositions*—to propositions supposed to be true, or taken as hypotheses—as well as to propositions that we take ourselves to know. Fully substantiating this claim would involve a long excursus into the epistemology of deductive inference, but consideration of a few cases may convince the reader that the claim is plausible. Suppose, for example, that my physicist informant does not know whether the body is accelerating, or whether a force is acting on it. All the same, being a competent physicist, he will still know that it is either not accelerating or has a force acting on it. How does he know that? We may grant for the sake of argument that he knows that the body is either accelerating or not. But how is that to yield knowledge that either a force is acting upon it or that it is not accelerating? In many circumstances, the only plausible answer will be this: because he is able to deduce that a force is acting on the body from the *supposition* that it is accelerating—where the deduction tracks the contextually relevant, physical consequence relation. Only having made that deduction from a supposition or hypothesis can he infer, in Russell's sense, from his knowledge that the body is either accelerating or not to attain knowledge that either it is not accelerating or has a force acting on it.

Now there is nothing at all alien in the idea that we may exercise our deductive capacities in reasoning from suppositions just as much as in reasoning from what we know. But reasoning from a supposition plainly demands a stronger condition for validity than Philonian consequence: the bare fact that either the conclusion is true or the premise is untrue is insufficient to underwrite the soundness of arguments from suppositions, for what is supposed to be the case may fail to be true. Let it be that the relevant body is not accelerating. In that case, any proposition whatever will be a Philonian consequence of "The body is accelerating." Yet the argument "Suppose the body is accelerating; then the moon is made of iron" is not a sound argument in physics. To explain why it is unsound, moreover, we have to invoke possibilities, or some cognate modal notion: the argument is unsound because it is physically possible for the body to accelerate without the moon's being made of iron. For this reason, Russell's strict Philonianism is in the end untenable. It relies on our knowing certain disjunctions, when the only plausible account of how we have that knowledge is through our applying our deductive capacities to suppositions, applications whose condition for validity cannot be Philonian.

Following Frege, Russell formalized his logic *more geometrico*, that is, as a system of putatively known axioms, and rules of inference by applying which a thinker can come to know theorems on the basis of those already known axioms. And, it might seem, the possibility of formalizing logic in this way shows that one could in principle avoid any reasoning from suppositions, and so rehabilitate Russell's Philonian account of consequence. But the availability of this style of formalization does not save the account. Formalizing logic in the geometric style is possible. But the question pressed on Russell was how the thinker comes by the relevant knowledge. If logic is formalized in the geometric style, the crucial questions become: how does a thinker know that the axioms are true, and how does he know that the rules preserve truth? And there is no plausible answer to these questions which does not appeal to the subject's ability to make deductions from suppositions. Certainly, the "elucidations" by which Frege originally justified his "geometrical" axioms and rules rely on that ability. Thus, in justifying *modus ponens*—from $A \rightarrow B$ and A , infer B —Frege reasons: "if B were not the True, then since A is the True, $A \rightarrow B$ would be the False" (Frege 1893, 25).¹³ In a case where *modus ponens* is applied in a Russellian inference, B will be true. So Frege's reasoning here traces the consequences of the false supposition that B is not true. His argument is none the worse for that, but it shows how even an adherent of a geometrical formalization of logic is driven to rely on our ability to reason from suppositions in grounding one of the formalization's primitive rules. And what goes for the rules goes equally for the axioms. If the geometrical style of formalization is now little more than a quaint anachronism, that is largely because it fails to show logical truths for what they are: simply the by-products of rules of inference that are applicable to suppositions—by-products that arise when all the suppositions on which a conclusion rests have been discharged.¹⁴

¹³ My translation; I have updated his logical symbolism.

¹⁴ Compare Michael Dummett: "The first to correct this distorted perspective, and to abandon the analogy between a formalization of logic and an axiomatic theory, was Gentzen ... In a sequent calculus or natural deduction formalization of logic, the recognition of statements as logically true does not occupy a central place ... The generation of logical truths is thus reduced to its proper, subsidiary, role, as a by-product, not the core, of logic" (Dummett 1981, 433–434).

It will pay dividends later to be more precise about the sort of supposing, or hypothesizing, that is relevant here. One pertinent division corresponds—roughly—to the division between indicative and counterfactual conditionals. “Suppose Shakespeare did not write *Hamlet*.” That sends us off in one direction: “In that case, Marlowe did; nobody else at the time could have done it.” Contrast “Suppose Shakespeare had not written *Hamlet*.” That sends us off along a different path, as it may be: “In that case, Stoppard would not have written *Rosencrantz and Guildenstern are Dead*.” A full account of the difference between these two kinds of supposing would be complicated, but the crux is that in elaborating suppositions of the first kind, but not the second, we are entitled to draw on what we know. We know that someone wrote *Hamlet*, and given that only Shakespeare and Marlowe were capable of writing it, the only consequent elaboration of the supposition that Shakespeare did not write it is that Marlowe did. In elaborating the supposition that Shakespeare had not written it, by contrast, our knowledge that someone did write it, while not lost, is temporarily set aside.

When I say that our capacity to elaborate suppositions will be part of any satisfactory account of a thinker’s logical knowledge, it is the first kind of supposition that I have in mind. Frege’s justification of *modus ponens* shows why it is this style of hypothesizing that is relevant. Prior to having made a Russellian inference from *A* to *B*, we shall not know whether *B* is true, so in entertaining the hypothesis “Suppose that *B* were not true,” we do not need to set any knowledge aside. What underpins our knowledge of logical truths, then, is our ability to elaborate suppositions or hypotheses as to what *is* the case; it is not essential that we should also be able to elaborate suppositions about what might have been the case.

5 Multiple conclusions

I suggested at the end of §2 that logic sets forth the general laws of consequence, where consequence is understood to relate some (possibly many) premises to a single conclusion. Some readers—multiple-conclusion logicians—will object to the restriction to single conclusions. The analysis of the previous section, I now argue, helps to justify that restriction.

We may agree that in some areas of our intellectual economy, multiple-conclusion consequence relations can play roles recognizably akin to those played by single-conclusion consequence relations. An important example is the way that consequence relations constrain combinations of acceptance and rejection of statements. The mark of an *attitude of acceptance*, as I shall use the term, is that a thinker who bears such an attitude to a proposition will thereby be mistaken unless the proposition to which he bears it is true. Truth is thus the norm for acceptance. Belief, plainly, is one such attitude, but there are others—notably the kind of “acceptance” which some philosophers of science say is the proper attitude to adopt towards currently well-confirmed scientific theories, given the strong “pessimistic” inductive evidence that those theories will eventually be refuted. It will be clear from this that the relevant notion of being mistaken is being mistaken as to the facts. A thinker may be mistaken in this sense even though his acceptance of a proposition is in no way irrational, nor even epistemically irresponsible. Similarly, the mark of an *attitude of rejection* is that a thinker who bears such an attitude to a proposition will thereby be mistaken unless the proposition to which he bears it is untrue. Just as belief is a paradigm attitude of acceptance, so disbelief is a paradigm attitude of rejection. And just as we express acceptance of a proposition by asserting it, so we express rejection of a proposition by denying it. This latter speech act may be performed, for example, by answering “no” to the corresponding yes-or-no question.

Part of the interest of single-conclusion consequence relations certainly lies in the way they bear on combinations of instances of acceptance and rejection. That bearing is partly captured in the following principle:

- (N) If a proposition *B* follows from a proposition *A*, then a thinker who accepts *A* and rejects *B* will be making at least one mistake as to the facts.

If *B* follows from *A*, then a certain combination of acceptance and rejection—viz. accepting *A* and rejecting *B*—will involve at least one mistake as to the facts. (It is a further question how one tries to correct this mistake.) Indeed, given the glosses placed on ‘accept’ and ‘reject,’ principle (N) admits of proof in a classical metalogic. For suppose that

B follows from A . Then we certainly have that if A is actually true then so is B . Given a classical metalogic, it follows from this that either A is not true or B is true. Now if A is not true, then any acceptance of A is mistaken. And if B is true then any rejection of B is mistaken. Either way, then, a thinker who both accepts A and rejects B will be making a mistake. But that is just what (N) says. This proof of (N), it may be noted, requires only that either B is actually true or A is actually untrue, given that B follows from A . It does not exploit any necessity there may be in the way B 's truth depends on A 's. So the proof of (N) goes through even for the weakest consequence relation—the relation of Philonian consequence, in which A stands to B except when A is actually true and B is actually untrue.¹⁵

Principle (N), and its proof, extends straightforwardly to the case of multiple premises. If B follows from $A_1, \dots, A_n$ then a thinker who accepts all the A_i and rejects B will be making at least one mistake as to the facts. What is striking, however, is that the principle, and the proof, also extend without strain to yield an account of the force of instances of multiple-conclusion consequence. Multiple-conclusion consequence obtains between two sets of propositions X and Y when it is impossible for all the members of X to be true without at least one member of Y 's being true. An argument parallel to the proof of (N) shows that if a set Y is a multiple-conclusion consequence of a set X then a thinker who accepts all the members of X while rejecting all the members of Y must be making at least one mistake.

This result has significance for the enterprise of multiple-conclusion logic. The rarity, to the point of extinction, of naturally occurring multiple-conclusion arguments has always been the reason why mainstream logicians have dismissed multiple-conclusion logic as little more than a curiosity (Tennant 1997, 320). And attempts by enthusiasts to alleviate the embarrassment here have often ended up compounding it. In

¹⁵ What of the converse of (N)? If, in the specified senses, it is a mistake to accept A and reject B , will B follow from A ? The answer is “no”; a version of Moore's paradox provides a counterexample. A thinker who accepts that it is raining but denies that he accepts that it is raining will be making a mistake as to the facts. But the proposition “He accepts that it is raining” is not a consequence of the proposition “It is raining.” It may not even be a Philonian consequence: we may suppose that it is in fact raining, but that no one accepts that it is. This sort of case is a problem for those who seek to explicate consequence in normative terms, but no such project is contemplated here.

the introduction to their textbook on the subject, Shoesmith and Smiley concede that multiple-conclusion proofs can scarcely be said to form part of the everyday repertoire of mathematics. "Perhaps the nearest one comes to them," they go on, "is in proof by cases, where one argues 'suppose $A_1 \dots$ then B , ... , suppose $A_m \dots$ then B ; but $A_1 \vee \dots \vee A_m$, so B .' A diagrammatic representation of this argument exhibits the downward branching which we shall see is typical of formalized multiple-conclusion proofs ... But the ordinary proof by cases is at best a degenerate form of multiple-conclusion argument, for the different conclusions are all the same (in our example they are all instances of the same formula B)" (Shoesmith and Smiley 1978, 4–5). "At best degenerate," though, hardly says it. I do not know how the word 'multiple' is used in Cambridge, but in the rest of the English-speaking world it is understood to mean 'more than one.' So an example of an argument in which all the conclusions (*sic*) are identical provides little justification for taking multiple-conclusion logic seriously. But since this is all that Shoesmith and Smiley provide by way of a positive case for deeming their system to be a branch of logic, readers of their book may be forgiven for closing it with a sigh on reaching page 5 of the introduction.

Our discussion suggests a better justification for their subject. Even if an instance of multiple-conclusion consequence yields no direct appraisal of any naturally occurring argument, it bears on the evaluation of certain combinations of acceptances and rejections in a way that recognizably generalizes (N). Seen from this angle, then, there seems to be no good reason to privilege multiple acceptances over multiple rejections. Let us grant that there are no naturally occurring multiple-conclusion arguments or proofs. If logic were primarily the theory of sound argument, or proof, this would matter. But, it might be claimed, its import lies in the way in which it constrains combinations of acceptances and rejections. On this view, there is no warrant for a lopsided restriction of the conclusion sets to singletons.

I think that this is the best case one can make for multiple-conclusion logic.¹⁶ Here, though, the best is not good enough.

¹⁶ Something like this case for multiple conclusions is advanced by Greg Restall (2005). But, whatever the fate of the objection to be pressed in the text, he overplays his hand in suggesting that 'Y is a multiple-conclusion consequence of X' can be explained as meaning 'The mental state of accepting all of X and rejecting all of Y would be self-defeating.' The mental state that consists of accepting that

The basic problem with the defense is that (*N*) does not capture anything like the full force of single-conclusion consequence. It says that when *B* follows from *A*, a thinker who accepts *A* and rejects *B* will be making a mistake. So he will. But—just to take one example among many—the principle says nothing about the case of someone who accepts *A*, who knows that *B* follows from *A*, but who refuses to accept *B*. A thinker who is in this position need not be making any mistake as to the facts. But the force of single-conclusion consequence is still eluding him. This is why we get irritated when we encounter such thinkers in our seminars. (“What do you mean, you refuse to accept *B*? You continue to adhere to *A*, and I have shown you that *B* follows from *A*.”) Yet this aspect of the force of consequence does not transfer to the multiple-conclusion case. A thinker who accepts all the statements in a set *X*, who knows that a set *Y* is a multiple-conclusion consequence of set *X*, but who refuses to accept any statement in *Y* need not be making any mistake. Of course, he will be making a mistake if he refuses to accept the claim that *some* member of *Y* is true. But that point is grist to the mill of skeptics about multiple-conclusion logic. Yet again, they will say, we can only understand an instance of multiple-conclusion consequence as an instance of single-conclusion consequence in which the conclusion is a disjunctive or existentially quantified claim.¹⁷

there will never be sufficient grounds for accepting or rejecting the proposition that there is a God, while rejecting the proposition that there is a God, is self-defeating. But “There is a God” is in no sense a consequence of “There will never be sufficient grounds for accepting or rejecting the proposition that there is a God.”

¹⁷ A cognate problem confronts Allan Gibbard’s attempt to extend the notion of validity to practical arguments—arguments whose conclusions are decisions rather than the simple acceptance of propositions. His leading example is Sherlock Holmes’s reasoning: “Either packing is now the thing to do, or by now it’s too late to catch the train anyway. It’s not even now too late to catch the train. Therefore packing is now the thing to do.” “We can see,” Gibbard writes, “what makes Holmes’s practical argument valid. It took the form: *F* or *P*, not *F*, therefore *P*. An argument of this form is valid, even if to accept *P* is to come a decision. To accept the premises and reject the conclusion would be to rule out every way that Holmes could become opinionated factually and fully decided in his hyperplan”—i.e. could come to have an opinion on every question of fact and to have made a decision about every question of what to do (Gibbard 2003, 59). Holmes would indeed be inconsistent if he accepted the practical argument’s premises and rejected its conclusion. But there remains a gap between rejecting the rejection of a practical conclusion and accepting it, so Gibbard does not capture the sense in which a successful deduction forces one to accept its conclusion.

The analysis of the previous section brings out, indeed, how impoverished a conception of the force of consequence (N) and its generalizations afford. Those generalizations record constraints on combinations of acceptances and rejections, and it is of some interest to have rules which yield such constraints. But constraints of that kind are only a small part of what makes single-conclusion logic important. Its rules are of interest chiefly because, by following them, a thinker can combine different sources of knowledge to come to know things that those sources do not establish by themselves. But this splicing together of different sources of knowledge to yield new propositional knowledge presumes many premises but only one conclusion. The epistemic importance of that splicing justifies logic's traditional focus on laws governing consequence relations that relate many premises to a single conclusion.

It is important to settle this matter, for the differences between multiple-conclusion and single-conclusion logical systems run deep. In particular, the pressure exerted by certain logical paradoxes is directed very differently in single-conclusion and multiple-conclusion frameworks. Ian Hacking has drawn attention to the "seemingly magical fact" that the same operational rules yield intuitionist logic in their single-conclusion form and classical logic in their multiple-conclusion form (Hacking 1979, 292–293). But this is just one illustration of the way the structural rules of a multiple-conclusion calculus bear more of the deductive weight than do the analogous rules in a single-conclusion system, and hence come more into contention in disputes between rival logical schools.

An even more striking illustration of this phenomenon concerns the multiple-conclusion cut rule, which Shoesmith and Smiley formulate as follows:

(*Multiple-conclusion cut*) If there exists a set of propositions Z such that $X, Z_1; Z_2, Y$ for each partition $\langle Z_1, Z_2 \rangle$ of Z , then $X; Y$. (Shoesmith and Smiley 1978, 29)

(A partition of Z is an ordered pair $\langle Z_1, Z_2 \rangle$ such that $Z_1 \cup Z_2 = Z$ and $Z_1 \cap Z_2 = \emptyset$.) Now in the envisaged justification of multiple-conclusion logic, all that mattered was that the relevant "logical" possibilities should include the actual circumstances. For that was enough to ensure that (N) and its generalizations are true. Accordingly, one of the

multiple-conclusion consequence relations to which the laws of multiple-conclusion logic apply must be the relation in which the only relevant possibilities are actual. This will be the material, or Philonian, multiple-conclusion consequence relation $\models$, which obtains between a set of premises X and a set of conclusions Y when either at least one member of X is not true or at least one member of Y is true. Thus the negation $\not\models$ of this relation will obtain between X and Y when each member of X is true but no member of Y is true. Instantiating the schematic ‘.’ in (*Multiple-conclusion cut*) with this Philonian relationship and contraposing, we reach

(*Philonian cut*) If $X \not\models Y$ then for every set Z of propositions, there exists a partition $\langle Z_1, Z_2 \rangle$ of Z such that $X, Z_1 \not\models Z_2, Y$.

The observation I wish to make is that (*Philonian cut*) contains a strong semantical assumption which goes far beyond the transitivity of consequence. This comes out most clearly if we apply the rule in a language with vague predicates: anyone who wishes to deny that vague predicates possess sharp boundaries or cut-off points is forced to deny (*Philonian cut*) and hence to deny the structural rule (*Multiple-conclusion cut*) of which it is a contraposed instance. To see this, consider a sequence $a_1, \dots, a_{100}$ of color samples which range gradually from a_1 , which is clearly red, to a_{100} , which is clearly orange and hence clearly not red. And consider the corresponding sequence of propositions $P_1, \dots, P_{100}$ in which P_i predicates redness of a_i . Let us then instantiate (*Philonian cut*) by taking the set X to be the singleton $\{P_1\}$ and Y to be the singleton $\{P_{100}\}$. The case has been constructed so that P_1 is true and P_{100} is not true, so we certainly have $P_1 \not\models P_{100}$. By (*Philonian cut*), then, we infer that

For every set Z of propositions, there exists a partition $\langle Z_1, Z_2 \rangle$ of Z such that $P_1, Z_1 \not\models Z_2, P_{100}$.

In particular, then, there must be a partition $\langle Z_1, Z_2 \rangle$ of the set $\{P_2, \dots, P_{99}\}$ such that $P_1, Z_1 \not\models Z_2, P_{100}$. That is to say, there must be a partition $\langle Z_1, Z_2 \rangle$ of the set $\{P_2, \dots, P_{99}\}$ such that all the members of Z_1 are actually true while none of the members of Z_2 is. Now clearly the only partition that could possibly have this property is

one where $Z_1 = \{P_2, \dots, P_k\}$ and $Z_2 = \{P_{k+1}, \dots, P_{99}\}$ for some integer k between 2 and 98. So we are committed to the existence of a k for which $P_1, \dots, P_k \not\models P_{k+1}, \dots, P_{100}$. But that is to say: we are committed to the existence of a k such that each proposition up to and including P_k in the sequence is true, while no proposition after P_k is true. We are committed, in other words, to a sharp cut-off point in the truth-values of the P_i s, and hence a sharp cut-off point in the redness of the a_i s. So anyone who wishes to deny that there is such a cut-off point must deny (*Multiple-conclusion cut*).

The moral I draw from this is that the content of (*Multiple-conclusion cut*) includes much more than the transitivity of consequence. The concentration of logical power in the structural rules is characteristic of multiple-conclusion systems, which may be likened to those high-powered sports cars whose engines need to be taken apart and rebuilt almost from scratch if one is to make even a minor adjustment to them: fun to drive, but a nightmare to maintain. But we have found reason to leave multiple-conclusion logics to the boy racers, and focus an inquiry into the nature of logic on the single-conclusion laws that sustain its ancient purpose of helping us to extend our knowledge of substantial, non-logical matters.

6 Broadly logical consequence regained

The argument of §4 showed that a consequence relation cannot be Philonian if it supports reasoning from a supposition that may be false. Even the consequence relation that underlies the tourist's argument supports such reasoning ("Suppose it's Tuesday; in that case, this must be Paris"), so even that relation involves some notion of possibility. Further reflection on the objection to Philonianism, moreover, suggests a principled ground for distinguishing a notion of broadly logical possibility and, *via* (*Cons*), a cognate relation of broadly logical consequence. As we have seen, some modal element will be implicit in any relation of consequence that is understood to relate premises that are entertained merely as suppositions to conclusions that are asserted only within the scope of those suppositions, as well as relating premises and conclusions that are asserted outright. All the same, most consequence relations are such that certain suppositions render them

inapplicable. When a classical physicist reasons “Suppose this body is accelerating; then a force is acting on it,” it may be reasonable to assess his reasoning against a relation of consequence, *R*, that excludes as impossible all circumstances that conflict with Newton’s laws of motion. Just for that reason, though, the argument “Suppose the body is accelerating; suppose also that Newton’s laws of motion are false; then a force is acting on the body” cannot sensibly be assessed against *R*. The introduction of the new suppositional premise creates a new argumentative context in which the contextually relevant consequence relation cannot be *R*.

The question then arises, whether there is a relation of consequence that is absolute in the sense of being applicable no matter what is supposed to be the case. Ian McFetridge proposed that logical consequence is such a relation, and that logical necessity is the correlative kind of necessity. He quotes with approval Mill’s dictum that “that which is necessary, that which *must* be, means that which will be, whatever supposition we make with regard to other things” (Mill 1891, book 3, chap. 5, sec. 6).¹⁸ And he goes on to suggest

that we treat as the manifestation of the belief that a mode of inference is logically necessarily truth-preserving, the preparedness to employ that mode of inference in reasoning from any set of suppositions whatsoever. Such a preparedness evinces the belief that, no matter what else was the case, the inferences would preserve truth ... A central point of interest in having such beliefs about logical necessity is to allow us to deploy principles of inference across the whole range of suppositions we might make. (McFetridge 1990, 153)

So a rule of inference is logically necessarily truth-preserving if it preserves truth, no matter what is supposed to be the case. And a truth will be logically necessary if it is true, no matter what is supposed to be the case. There is no suggestion that the only rules that are logically necessarily truth-preserving are the formal rules presented in logic books. For example, we are surely prepared to apply the rule of infer-

¹⁸ Not chapter 4, as the editors of McFetridge’s *Nachlass* claim. Mill was characterizing causal necessity.

ence “From ‘ x is red,’ infer ‘ x is colored’” in reasoning from any set of suppositions whatsoever. Thus we may read McFettridge as advancing an account of the Moorean, “broad” notion of logical necessity that we are seeking to elucidate.

From the perspective recommended here, McFettridge’s account of the notion is attractive. The thesis that logical necessity is truth no matter what is supposed to be the case coheres nicely with the original ground for recognizing a modal element in consequence—namely, its applicability to assessing reasoning from suppositions. And the idea that logical consequence is distinguished from other consequence relations through its applicability to any suppositions whatever is supported when we reflect on how we trace the consequences of a supposition that is contrary to an accepted logical law. A classical logician who reduces to absurdity the supposition that $\neg(P \vee \neg P)$ need not consider how his logic might change if excluded middle had a false instance. In his *reductio*, he will apply the normal classical logical rules without demur. I do not mean to imply that the rules of classical logic cannot be challenged: they can be. But someone who accepts classical logic will apply its rules to all suppositions whatever, even to suppositions that are contrary to classical logical laws.

For these reasons, I recommend accepting McFettridge’s account of broadly logical consequence and broadly logical necessity. Apart from one class of apparent exceptions, which I shall analyze in detail below, his account delivers the expected classifications when applied to paradigms of arguments that are, and that are not, logically valid in the broad sense. More importantly, the relation he identifies as broadly logical consequence is theoretically significant: since any consequence relation has a modal aspect by virtue of being applicable to suppositions, special interest naturally attaches to a relation that is applicable to absolutely any supposition. We need to inquire, though, how the cognate notion of logical necessity relates to other central modal notions.

There seems to be no reason to suppose that every proposition that is logically necessary in the present sense is knowable *a priori*. Indeed, there is no general reason to suppose that every logically necessary proposition is knowable *tout court*. If a proposition is logically necessary, it will be true no matter what is supposed to be the case. But its having that property does not imply that it is possible to know it.

In denying that logical necessity entails knowability *a priori*, I am at odds with Dorothy Edgington (2004). Pre-Kripkean discussions of validity, she remarks, made us “familiar with two thoughts: first, an argument is valid if and only if it is necessary that the conclusion is true if the premises are true; and second, if an argument is valid, and you accept that the premises are true, you need no further empirical information to enable you to recognize that the conclusion is true ... Given Kripke’s work, and taking ‘necessary’ in its metaphysical sense, these two thoughts are not equivalent” (9). So we have to choose what the criterion for validity is to be. It is, she claims, “the least departure from traditional, pre-Kripkean thinking, and more consonant with the point of distinguishing valid from invalid arguments, to take validity to be governed by epistemic necessity, i.e. an argument is valid if and only if there is an *a priori* route from premises to conclusion” (10).

I agree that an argument’s broadly logical validity does not consist in its premises’ metaphysically necessitating its conclusion, but Edgington’s positive account does not, I think, capture the concept we are trying to explicate. First, there are propositions which we can know *a priori* but which we are reluctant to classify as logically true even on the most generous demarcation of the bounds of logic. Some people now know *a priori* that when the index n is greater than two, there are no integral solutions of the equation $x^n + y^n = z^n$. All the same, it is not simple prejudice to resist the claim that Fermat’s Last Theorem is *logically* true. The ground for resistance is not the complexity of the proof: there are long and complex logical deductions. Rather, it is the heavy ontological and ideological commitments of the mathematical theories on which the proof depends. There are good reasons for postulating the truth of those theories: if there were not, then we should not have a proof of the theorem. But the postulates lie far beyond anything that is required for the regulation of *reasoning* (as opposed to calculating, and then theorizing about the numbers invoked in making calculations ...), and hence lie outside the widest plausible bounds of logic.

Perhaps Edgington would respond to this point by emending her position, and proposing that an argument is broadly logically valid if and only if there is a route from its premises to its conclusion that a thinker may traverse purely by exercising his ability to reason. But this emended theory still faces a second objection. On the emended proposal, the conclusion of a valid argument must be deducible from its

premises, but while this principle may have been part of “traditional thinking about validity,” it is surely too strong. It excludes notions of validity (such as validity in full second-order logic) for which no complete set of deductive rules can be given. A conclusion may be a second-order consequence of some premises even when it cannot be deduced from those premises. Inasmuch as old-fashioned thinking about validity overlooked this point, it was simply wrong. As for the suggestion that her account comports best with the point of distinguishing valid from invalid arguments, many philosophers have taken broadly logical consequence to be a relation to which our rules of deduction must *answer*: a rule is to be rejected as (logically) unsound if it enables us to deduce from some premises a purported conclusion that does not follow from them in the broad sense. But the idea that deductions answer to consequence would become incomprehensible if consequence were taken to consist in deducibility.

So much for apriority. How does the present notion of logical necessity relate to Kripkean metaphysical necessity?¹⁹ It is reasonably clear that some metaphysical necessities are not logically necessary. It is metaphysically necessary that Hesperus is identical with Phosphorus (if Hesperus exists). It is tolerably clear, however, that this identity is not logically necessary in McFetridge’s sense. A thinker ignorant of elementary astronomy can suppose that Hesperus is distinct from Phosphorus, without setting any knowledge aside. Under that supposition, the statement “Hesperus is identical with Phosphorus” will be false. So the statement is not true no matter what is supposed to be the case.

What of the converse, though? Are there logical necessities that are not metaphysically necessary? McFetridge claims not. Indeed, he advances an interesting general argument that purports to show that

¹⁹ I cannot discuss the notion of metaphysical modality in any detail here, but I assume that a metaphysical possibility must respect the actual identities of things—in a broad sense of ‘thing’ that encompasses stuffs such as water, and phenomena such as heat, as well as individual objects such as the planet Hesperus. This, at least, seems to be the gloss that best vindicates Kripke’s attributions of metaphysical necessity—notably his claims that ‘Water is H₂O’ and ‘Heat is the motion of molecules’ express metaphysically necessary propositions (1980, 99, 128–133), and his crucial thesis that if $x=y$ then it is metaphysically necessary that $x=y$ (1980, 3–5, 97–105). For an elaboration and defense of this gloss, see Fine 2002.

logical necessity (conceived as he conceives it) is the strongest form of non-epistemic necessity.²⁰ His argument rests on two assumptions. First, “that adding extra premises to a [logically] valid argument cannot destroy its validity ... If the argument ‘ P ; so Q ’ is valid then so is the argument ‘ P, R ; so Q ’ for any R .” Second, “that there is this connection between deducing Q from P and asserting a conditional: that on the basis of a deduction of Q from P one is entitled to assert the conditional, *indicative or subjunctive*, if P then Q ” (McFetridge 1990, 138; emphasis in the original). The argument then runs as follows. Suppose it is logically necessary that if P then Q . Suppose also, for *reductio*, that in some other sense of ‘necessary,’ it is not necessary that if P then Q . Then, in the sense of ‘possible’ that corresponds to this other sense of ‘necessary,’ it is possible that P and not Q . But

if that *is* a possibility, we ought to be able to describe the circumstances in which it would be realized: let them be described by R . Consider now the argument “ P and R ; so Q .” By the first assumption if “ P ; so Q ” is valid, so is “ P and R ; so Q .” But then, by the second assumption, we should be entitled to assert: if P and R were the case then Q would be the case. But how can this be assertible? For R was chosen to describe possible circumstances in which P and not Q . I think we should conclude that we cannot allow, where there is such an R , that an argument is valid. (138–139)

When it is logically necessary that if P then Q , however, the argument “ P ; so Q ” will be valid. So in that case there is no such R . So it is in no sense possible that P and not Q . So it is in every sense necessary that if P then Q . Hence McFetridge’s conclusion: “logical necessity, if there is such a thing, is the highest grade of necessity.”

What should we make of this argument? Given McFetridge’s conception of broadly logical consequence, his first assumption is unsailable. If “ P ; so Q ” is logically valid, then the inference from P to Q

²⁰ Thus he expressly excludes the notion of necessity that corresponds to “mere time- and person-relative epistemic possibility—which can be asserted even when logical possibility cannot. I mean the notion expressed by that use of ‘It may be that p ’ which just comes to ‘For all I know, not p ’” (McFetridge 1990, 137). I shall follow him in setting these cases aside.

preserves truth no matter what is supposed to be the case. In particular, then, it preserves truth if R is supposed to be the case. So the argument " P, R ; so Q " will also be valid.

As for the second assumption, the part claiming that we may assert the indicative conditional "If P then Q " on the strength of a deduction of Q from P is an application of the rule of conditional proof, a rule that is often held to specify the sense of the indicative conditional. The corresponding claim for subjunctives is less obvious, and I shall eventually suggest that it needs to be qualified, but we should still acknowledge that even the subjunctive part of McFetridge's second assumption is very plausible. As it concerns subjunctives, the second assumption amounts to this: that we can apply our capacity for broadly logical deduction in elaborating counterfactual suppositions. And the worry about rejecting this assumption is that if we were not able to apply *that* capacity, we should be quite unable to elaborate counterfactual suppositions at all. Of course, in elaborating a given counterfactual supposition, some of our deductive capacities will not be applicable. A capacity for deducing the consequences of suppositions according to the principles of classical physics, for example, is quite inapplicable in elaborating the counterfactual supposition "Suppose that the gravitational force between two bodies had varied with the inverse cube of the distance between them." Just for that reason, though, we badly need some rules which are guaranteed to yield consequent elaborations of our counterfactual suppositions. Since logic (broadly conceived) is traditionally supposed to apply to anything that is so much as thinkable, one would expect logical rules (even rules that are logically valid in the broad sense) to provide what we need. If they do not, it is wholly unclear what else does, or could, provide what is needed in elaborating a counterfactual.

All the same, the two assumptions have some unpalatable consequences. In support of her positive account of broadly logical validity, Edgington elaborates Kripke's example of Leverrier, who postulated a nearby, hitherto unobserved planet as the cause of certain observed perturbations in the orbit of Uranus, and who introduced the name 'Neptune' as a term which was to stand (rigidly) for such a planet, if indeed there was one (see Kripke 1980, 79). About this case, she remarks that "it is epistemically possible that [Leverrier's] hypothesis was wrong—that there is no such planet. But if his hypothesis is

right—if Neptune exists—it is the planet causing these perturbations. And this conditional is known *a priori*—at least by Leverrier: it follows from his stipulation about the use of ‘Neptune’” (Edgington 2004, 7). Thus Leverrier’s “argument from the premise that Neptune exists to the conclusion that it causes these perturbations is trivial”—and hence valid. Although I have denied that validity follows from the existence of an *a priori* route from premises to conclusion, Edgington’s assessment of this particular argument—call it (*E*)—still seems right. Although it is not formally valid, the triviality of the inference from premise to conclusion means that (*E*) is valid in the broad sense we are trying to elucidate, so that its premise broadly logically necessitates its conclusion.

On McFetridge’s principles, though, it seems that (*E*) cannot be valid. Certainly, he thought he was committed to denying its validity. Discussing an ancestor of the paper of Edgington’s from which I have been quoting, he writes:

Following Kripke and Evans Edgington claims, and I agree, that [Leverrier] knows *a priori* that if Neptune exists it is a planet causing such and such perturbations. Thus, on her account, the argument: “Neptune exists, so Neptune causes such and such perturbations” is deductively valid. But there certainly is a “timeless” metaphysical possibility that the premise should have been true and the conclusion false: suppose Neptune had been knocked off course a million years ago. What then, of the argument: “Neptune exists and was knocked off course a million years ago, so Neptune is the cause of these perturbations”? If the original argument is valid so is this one (by the first assumption). But if it is we ought (by the second assumption) to be entitled to assert: if Neptune had existed and been knocked off its course a million years ago then it would have been the cause of these perturbations. But of course we are not entitled to assert that: had the antecedent been true the consequent would have been false. (McFetridge 1990, 139)

McFetridge’s two assumptions, then, appear to reduce to absurdity the claim that argument (*E*) is broadly logically valid. However, the inference in (*E*) is trivial, and it seems hard to allow that a trivial inference

is not broadly logically valid. Something in the analysis, then, seems to have gone seriously wrong. But what could it be?

The problem is not confined to this one example. As Edgington recognizes, parallel cases may be constructed whenever we have what Gareth Evans called a “descriptive name.” Evans’s own example was ‘Julius,’ which he introduced as a descriptive name that rigidly designates the person (if there was one) who actually invented the zip fastener (Evans 1979). Thus Edgington invites us to “consider the argument [*F*]: Julius was a mathematician; the person who invented the zip fastener emigrated to Tahiti; therefore, some mathematician emigrated to Tahiti.” As before, the triviality of the inference here makes it natural to classify (*F*) as valid in the broad sense. “Yet there are metaphysically possible situations in which the premises are true and the conclusion false, namely, ones in which Julius, the actual inventor of the zip fastener, did not do so and someone else, who emigrated to Tahiti, did, and no mathematician emigrated to Tahiti” (2004, 9–10). Contrary to McFetridge’s master thesis, then, we seem to have a case of a logical necessary proposition that is not metaphysically necessary.

Edgington tries to resolve the difficulty by rejecting the subjunctive part of McFetridge’s second assumption. “We are familiar with the fact that an indicative and a subjunctive ‘If *A*, *B*’ can disagree,” she says. “In the indicative, the antecedent presents something as an epistemic possibility, while in the subjunctive the antecedent typically presents something as *not* an epistemic possibility, but as something which *was* a real possibility. Each kind of conditional goes with a different kind of possibility. McFetridge’s second assumption, that there is a unitary sense of ‘possible’ that governs both, is not obligatory” (13). But his second assumption does not require a unitary space of possibilities. All it requires is that broadly logical deduction should be applicable in elaborating both epistemic possibilities and the “once real” possibilities that Edgington takes subjunctives to present. Edgington does not explain why broadly logical deduction should be inapplicable to the latter cases, nor does she tell us what inferential principles we can rely on if it is not so applicable. In the absence of such an explanation, we are left with an *aporia*, perhaps even with a paradox.

7 The paradox resolved

I think we can resolve the difficulty here by attending more closely to some peculiar features of our problem cases.

Edgington assesses argument (*E*) as valid, and that assessment is plausible. But it is only plausible in a context where it is common knowledge that the name ‘Neptune’ was introduced by Leverrier’s stipulation, and where people still understand the name in strict conformity with that stipulation. While the very earliest astronomical uses of the name may have conformed to this requirement, later ones do not: within a few months of Leverrier’s having made his conjecture, Neptune had been sighted through telescopes and people came to understand ‘Neptune’ in ways that did not depend on knowing how the name had originally been introduced. Were argument (*E*) to come from such a person—for example, from someone who understands the name as standing for the eighth most distant planet from the sun—we would have no inclination at all to classify the argument as valid. In other words, we shall count (*E*) as valid only in contexts where the component occurrence of ‘Neptune’ is understood to be a descriptive name for the planet (if there is one) that is the cause of certain observed perturbations in the orbit of Uranus. As for the argument (*F*), it is already explicit that ‘Julius,’ as it occurs there, is to be understood as a descriptive name.

Now descriptive names have some rather special semantical features. They are rigid, in the sense that the truth or falsity with respect to a counterfactual situation of a predication involving one of them depends on the properties, in that situation, of the descriptive name’s actual referent.²¹ Indeed, ‘Neptune’ must be treated as rigid in this sense if the conditional “If Neptune exists, then some planet is the cause of these perturbations” is to have a false necessitation. On the other hand, they differ from ordinary proper names inasmuch as it is not constitutive of a descriptive name’s meaning that it should have the reference that it actually has. What sustains the use of the ordinary proper name ‘the Moon’ among English speakers is their common knowledge, concerning the Moon, that the name designates it. By contrast, what

²¹ For an argument that the concept of rigid designation is best understood along lines such as these, see Cartwright 1997.

sustained the use of name 'Neptune,' during the short period when it was a descriptive name, was not speakers' common knowledge, concerning an object that is in fact Neptune, that the name designates it. Rather, it was their common knowledge that 'Neptune' designates whatever planet causes the observed perturbations, if there is one such planet, and otherwise designates nothing. Similarly, what sustains the current use of the name 'Julius' among philosophical logicians is their common knowledge that the name designates whoever invented the zip, if there was one such person, and otherwise designates nothing. Let us take the meaning of an expression, in a given use, to be given by the proposition, common knowledge of which among the relevant speakers sustains that use. Then the meaning of 'Neptune' (during that early period of its use within astronomy) will consist in its designating any planet that was the unique planetary cause of the observed perturbations; similarly, the meaning of the name 'Julius,' in its current use among philosophers, will consist in its designating any person who uniquely invented the zip.

This unusual combination of semantic features has some interesting consequences. As we have seen, the rigidity of the name 'Neptune' means that the conditional (ϵ) "If Neptune exists, then some planet is the cause of these perturbations" has a false necessitation: because 'Neptune' is taken to designate its actual referent, even when we are evaluating propositions involving modal operators, the proposition "It is possible that Neptune should have existed while a comet caused the perturbations" is true. On the other hand, our thesis about the meaning of 'Neptune' implies that there is no possible circumstance in which (ϵ) is false, meaning what it actually means. For in any possible circumstance in which 'Neptune' and 'exists' mean what they actually mean, the statement "Neptune exists" will be true only if there is a planet which is the cause of the relevant perturbations, and in any such circumstance the conditional's consequent will be true, assuming that it too means what it actually means. Despite its having a false necessitation, then, it is necessary that conditional (ϵ) is true, when taken in its actual sense. The same is true of "If Julius was a mathematician and the person who invented the zip fastener emigrated to Tahiti, then some mathematician emigrated to Tahiti." For the reason Edgington gives, the necessitation of this conditional is false. But there is no possible circumstance in which the conditional is false, meaning what it actually means.

This shows that in analyzing propositions involving descriptive names, we need to distinguish two alethic relations between propositions and possibilities. Suppose that a proposition *A* says that *P*. Then we can explain as follows the notion of *A*'s being true *with respect to* a possibility *x*: were *x* to obtain (or had *x* obtained), it would be (or would have been) the case that *P*. We should contrast this with the notion of *A*'s being true *in* a possibility *x*: were *x* to obtain (or had *x* obtained), *A* would be true (or would have been true), meaning what it actually means. When a proposition contains a descriptive name, the notions are liable to come apart. The zip fastener was in fact invented by an American, who was known to his friends as 'Whitcomb L. Judson.' Let *x* be a possibility in which the zip was not invented by Judson but by a Russian, but in which Judson retains his actual nationality. Since 'Julius' rigidly designates Judson, it would be wrong to say: "Had *x* obtained, it would have been the case that Julius was Russian." So "Julius was Russian" is not true with respect to *x*. But we may say: "Had *x* obtained, the proposition 'Julius was Russian' would have been true, meaning what it actually means." For had *x* obtained, the name 'Julius' (meaning what it actually means) would have designated a Russian. Thus "Julius was Russian" is true in *x*.

These two alethic relations between a proposition and a possibility correspond to two senses in which a proposition might be said to be necessary. Again suppose that the proposition *A* says that *P*. Suppose too that *A* is true with respect to every possibility. Then, no matter what were to obtain, or what had obtained, it would be, or would have been, the case that *P*. Thus it will be necessarily the case that *P*. The converse implication also holds, so we may say that the proposition *A* is true with respect to every possibility if, and only if, it is necessarily the case that *P*. In other words, a proposition is true with respect to every possibility just in case it *has a true necessitation*. Following Evans, let us call such a proposition *superficially* necessary. By contrast, if *A* is true in every possibility then, no matter what were to obtain or had obtained, *A* would be or would have been true, meaning what it actually means. Thus a proposition is true in every possibility just in case it is *necessarily true*, meaning what it actually means. We may call such a proposition *deeply* necessary.

This distinction applies to each underlying species of necessity. Thus a proposition is superficially logically (metaphysically, physically)

necessary just in case it has a true logical (metaphysical, physical) necessitation, which obtains just in case it is true with respect to every logical (metaphysical, physical) possibility. And a proposition is deeply logically (...) necessary just in case it is logically (...) necessarily true (meaning what it actually means), which obtains just in case it is true in every logical (...) possibility. Again, propositions involving descriptive names provide examples where these distinctions mark a difference. Because Judson need not have invented the zip, the proposition “If anyone invented the zip, Julius did” is superficially logically contingent: its logical necessitation is false. But it is logically necessarily true, taken in its actual meaning, so it is deeply logically necessary. Because the name ‘Julius’ is rigid, the statement “Julius is identical with Judson” has a metaphysically true necessitation: Julius could not have been anyone other than Judson. Thus it is superficially metaphysically necessary. But the proposition is not deeply metaphysically necessary: it is not metaphysically necessary that it is true, meaning what it actually means. The descriptive name ‘Julius’ could have designated someone other than Judson without any change in its meaning—that is, without any change in the facts, common knowledge of which sustains the relevant use of the name.

Crucially for present purposes, there is a corresponding bifurcation in the sense of the phrase ‘an argument’s premises necessitate its conclusion.’ By this one could mean: the conclusion is true with respect to any possibility with respect to which all the premises are true. When this obtains, let us say that the premises *superficially* necessitate the conclusion. Let us define an argument’s *validating conditional* to be the indicative conditional whose antecedent is the conjunction of the argument’s premises, and whose consequent is its conclusion; thus (ϵ) above is the validating conditional for the argument (E). Then an argument’s premises will superficially necessitate its conclusion if and only if the argument’s validating conditional has a true necessitation. A second thing one might mean by ‘The premises necessitate the conclusion’ is that the conclusion is true *in* any possibility in which all the premises are true. When this second condition obtains, let us say that the premises *deeply* necessitate the conclusion. An argument’s premises will deeply necessitate its conclusion if and only if it is necessary that the argument’s conclusion is true if its premises are true (meaning what they actually mean). This will obtain if and only if the argument’s

validating conditional is necessarily true (meaning what it actually means). As before, we can and should make this distinction in respect of each species of necessity.

I think that these distinctions give us what we need to resolve the paradox presented by McFetridge's argument that logical necessity is the strongest form of necessity.

When an argument is logically valid, its premises logically necessitate its conclusion. But do they necessitate it superficially or deeply? When the argument is " P ; so Q ," does validity imply that it is logically necessary that Q if P , or does it imply that it is logically necessary that ' Q ' is true if ' P ' is true (meaning that they actually mean)? In addressing this question, it does not help to ask which answer is closer to the traditional explanation of consequence or validity, for there is no reason to think that those who framed that explanation had considered the comparatively rare cases where the answers diverge. But when we reflect on why we are interested in logical validity, and on the reasons for recognizing a modal element in consequence, I think it becomes clear that validity requires that the premises should deeply necessitate the conclusion. I shall argue, in other words, that the principle (*Cons*) of §3 was formulated correctly: consequence obtains when the conclusion is true *in every possibility* in which all the premises are true.

Why should this be? We are interested in deduction, I argued, primarily because it affords a means of extending our knowledge of premises to knowledge of a conclusion. And we are interested specifically in logical deduction because it affords such a means, no matter what we suppose to be the case. As I remarked at the end of §4, however, the relevant kind of supposing is supposing that such-and-such is the case, not supposing that such-and-such might have been the case. In considering what might have been the case, we shall in general set some things we know aside, and we shall not in general wish to do this when trying to extend our knowledge. Now one thing a competent thinker will certainly know is what his words mean, so this is part of the background knowledge that a thinker may be assumed to possess; moreover, this knowledge will not be set aside as the thinker elaborates suppositions as to what is the case. In a given case, it will be possible to state this knowledge in the material mode. Thus someone who has heard Leverrier's stipulation will know that Neptune, if it exists at all, is the planet that causes certain observed perturbations in the orbit of

Uranus. But in stating a general principle about validity we can hardly do otherwise than to make a semantic ascent and say: if an argument is valid, then no matter what supposition we make that is consistent with the component terms' meaning what they actually mean, if the premises are true then the conclusion is true. But that just is to say that validity requires that the premises should *deeply* necessitate the conclusion.

With this point settled, let us scrutinize again McFettridge's argument that logical necessity is the strongest form of necessity. For the reasons given in the previous section, we should certainly accept its first premise: if the argument " P ; so Q " is valid then so is the argument " P, R ; so Q " for any R . We may also accept the part of its second premise that concerns indicatives: on the strength of a deduction of Q from P , one is entitled to assert the indicative conditional, "If P then Q ." The argument relies, though, on the corresponding claim for subjunctive or counterfactual conditionals; is that claim true? Because a valid argument's premises deeply necessitate its conclusion, it will be logically necessary that the conclusion is true if the premises are, so we are entitled to assert the subjunctive or counterfactual conditionals "Were the premise true (meaning what it means), the conclusion would be true (meaning what it means)" or (with the same provisos) "Had the premise been true, the conclusion would have been true." But the use of the truth predicate in these subjunctive conditionals is essential: where the premise of the argument is that P , and the conclusion is that Q , we cannot always "disquote" and assert "Had it been that P , it would have been that Q ." Those habituated to think that the truth predicate is invariably redundant may find this surprising, but again the Neptune example brings out the difference clearly. "If Neptune had existed and been knocked off its course a million years ago, then it would have been the cause of these perturbations" is straightforwardly false. But if we take a counterfactual conditional with an impossible antecedent to be vacuously true, then it *is* true to say: "If the proposition 'Neptune exists, and was knocked off its course a million years ago' had been true (meaning what it actually means), then the proposition 'Neptune is the cause of these perturbations' would have been true (meaning what it actually means)." For it is impossible for the proposition "Neptune exists, and was knocked off its course a million years ago" to be true while meaning what it actually means.

All the same, the qualified version of McFetridge's second premise that we have allowed to be true suffices for his argument to reach its intended conclusion. Given that P logically entails Q , the argument proceeds by reducing to absurdity the supposition that in some sense of 'possible,' it is possible that P and not Q . And we can emend the crucial passage in McFetridge as follows:

if it is a possibility that P and not Q , we ought to be able to describe the circumstances in which it would be realized: let them be described by R . Consider now the argument " P and R ; so Q ." By the first assumption if " P ; so Q " is valid, so is " P and R ; so Q ." But then, by the second assumption, we should be entitled to assert: if P and R were true (meaning what they mean) then Q would be true (meaning what it means). But how can this be assertible? For R was chosen to describe possible circumstances in which P and not Q , so if R were true, then "not Q " would be true. I think we should conclude that we cannot allow, where there is such an R , that an argument is valid. (Compare McFetridge 1990, 138–139)

The subsidiary conclusion is still absurd, so taking proper care over semantic ascent does not stop McFetridge from achieving his *reductio*. So we should accept his conclusion that any (non-epistemic) possibility is a logical possibility.

Care over semantic ascent and descent, though, is precisely what is needed to reconcile this conclusion with the intuitive validity of the problematic argument (E). As we saw, it is plausible to classify (E) as valid in the broad sense; our paradox arose because McFetridge seemed to be forced to reject that classification. If (E) were valid, he thought, then on his own principles a speaker would be entitled to assert the subjunctive conditional "If Neptune had existed and been knocked off its course a million years ago then it would have been the cause of these perturbations," when patently this is not assertible. We now see, though, that the assertibility of this subjunctive conditional does not follow from (E)'s validity. Because (E) is valid, so must be the argument: "Neptune exists and was knocked off course a million years ago, so Neptune is the cause of these perturbations." The validating conditional of this latter argument must then be deeply logically necessary; that is to say, the validating conditional is true *in every*

possibility. And indeed it *is* true, vacuously, in every possibility: there is no possibility *in* which its antecedent—viz., the proposition “Neptune exists and was knocked off course a million years ago”—is true. But that does not mean that the validating conditional is true *with respect to* every possibility, which is what is required for it to have a true necessitation, and hence for our subjunctive conditional to be assertible. Once the appropriate distinctions are drawn, then, we see that (*E*)’s validity does not commit one to the assertibility of the unacceptable subjunctive conditional. The sense of paradox is finally dispelled.

Because the analysis has been rather intricate, a summary may help:

(1) We should accept McFetridge’s thesis that logical necessity is the strongest form of non-epistemic necessity. Any non-epistemic possibility is a logical possibility. Moreover, McFetridge’s argument for his thesis is essentially correct.

(2) The problematic argument (*E*) does not threaten this thesis by providing an example of a situation which is metaphysically possible but logically impossible. Rather, it shows the need, when analyzing arguments involving descriptive names, to distinguish between truth in a possibility and truth with respect to a possibility. Since (*E*) is valid, its conclusion is true in any logical possibility in which its premise is true. Since any metaphysical possibility is a logical possibility, (*E*)’s conclusion will be true in any metaphysical possibility in which its premise is true. The circumstance that McFetridge describes, in which Neptune was knocked off its actual course a million years ago, is metaphysically possible, and hence logically possible. But it is not a possibility in which the premise of (*E*) is true and the conclusion false. Rather, it is a possibility *with respect to* which the premise of (*E*) is true and the conclusion is false. Thus both the metaphysical necessitation, and the logical necessitation, of (*E*)’s validating conditional are false. This explains why we cannot correctly assert the subjunctive conditional: “If Neptune had existed and been knocked off its course a million years ago, then it would have been the cause of these perturbations.”

(3) The distinction between truth in a possibility and truth with respect to a possibility, and the cognate distinction between deep and superficial necessitation, can be made out only if we recognize that certain disquotational principles are not necessarily true. Of course,

no one thinks it necessary that the words ‘the Moon,’ identified simply as a linguistic type, should designate the Moon. The words, identified purely typographically, could always have been used to stand for something else. But it is plausible to hold it to be necessary that the words ‘the Moon,’ used as they are actually used in English, should designate the Moon. For it is constitutive of that use of the words that they should designate the Moon. If we think of the English language as constituted by its semantics, as well as by its syntax, phonology, etc., we can put this by saying that it is necessary that ‘the Moon’ should designate the Moon *in English*.²² But while the like will hold good of other ordinary proper names, it will not hold good of descriptive names. It is not necessary, for example, that the name ‘Julius’ should designate Julius in English—or better, in that dialect of English common among philosophical logicians. Since ‘Julius’ is a rigid designator, if this were necessary it would follow that Whitcomb L. Judson is necessarily such that the name ‘Julius’ designates *him* in English. But the descriptive name ‘Julius’ can mean what it actually means while standing for someone else. The use of the term ‘meaning’ that yields this result is in no way capricious. The underlying principle is that an expression’s meaning is given by the proposition, common knowledge of which sustains the expression’s use. The relevant proposition in the present case is simply that ‘Julius’ designates whoever uniquely invented the zip.

(4) The analysis does imply some restrictions on the use of our logical capacity in elaborating counterfactual suppositions—suppositions that such-and-such might have been the case. Supposing that something might have been the case in general involves setting some knowledge aside; and sometimes what is set aside will include the knowledge that sustains an expression’s use (the knowledge that gives its meaning, as I use that term). Thus in supposing that Neptune was knocked far off course a million years ago, we set aside our knowledge that it was around to perturb Uranus in the late nineteenth century, which is the knowledge that sustained the early use of the name ‘Neptune.’ In these cases, a supposition’s broadly logical consequences will not be part of an elaboration of it. It would be absurd to say “Neptune

²² In my paper (Rumfitt 2001), I called languages individuated in this way “Peacockian languages,” after Peacocke 1978.

might have been knocked off course a million years ago; and in that case Uranus would also have been knocked off course, so that its orbit could have been perturbed by Neptune.” But in practice we have no difficulty avoiding the absurdities, and our analysis explains why they are to be avoided.

I have been defending McFetridge’s claim that logical necessity implies metaphysical necessity. Does the distinction between truth *in*, and truth with respect to a possibility explain, or explain away, every apparent instance of a proposition’s being logically necessary but not metaphysically necessary—even when the proposition contains no descriptive name? I cannot canvass here all the putative examples that philosophers have discussed, but a rather different sort of case is “A red object looks red to a normal viewer in optimal viewing conditions,” where a “normal” viewer is understood to be one who is neither blind, nor color-blind, nor otherwise unable to identify or discriminate between colors by sight. It is plausible to claim that this proposition is (broadly) logically necessary. From the supposition that an object is red, one may infer that it looks red to a normal viewer in optimal viewing conditions, no matter what else one supposes to be the case. But it is equally plausible to deem the proposition to be metaphysically contingent, in the sense of having a false metaphysical necessitation. Human beings could have been constituted so as to see red things as violet, violet things as red, and so on, in which case red things would not have looked red, even in optimal viewing conditions, to the people who are best able to identify colors by sight.

Even though the present proposition contains no descriptive name, the distinction between truth *in* and truth with respect to a possibility is still of use in analyzing it, and in heading off any threat it may present of a renewed paradox. The possibility described in the last sentence of the previous paragraph is one with respect to which our proposition is false: the proposition has a false (metaphysical) necessitation. But there are no possibilities *in* which it is false. That is why the inference “Such-and-such is red; so such-and-such looks red to a normal viewer in optimal viewing conditions” is intuitively valid. We could confirm this by spelling out the knowledge that sustains the use of the expressions ‘red,’ ‘looks red,’ ‘normal viewer,’ and so on. Doing this properly would involve giving a theory of meaning for color terms, something I shall not essay here. But it is a plausible constraint on any

such theory that it should entail that, whenever the words ‘red,’ ‘looks red,’ ‘normal viewer,’ and so on, have their actual meanings, ‘red’ is co-extensive with the expression ‘looks red to a normal viewer in optimal viewing conditions.’ If that constraint is met, then the proposition will be true in any possibility in which it means what it actually means, so the proposition is deeply logically necessary and hence deeply metaphysically necessary, even though it is superficially metaphysically contingent, and hence superficially logically contingent. Paradox is averted, then, essentially as before.

If space permitted, it would be worth comparing this “metalinguistic” account of the distinction between deep and superficial varieties of necessity with the “two-dimensional” account, first proposed by Martin Davies and Lloyd Humberstone (1980), which has come to predominate.²³ Independently of that comparison, though, I hope to have shown how McFetridge’s account of logical necessity can be defended against the paradox that arguments such as (E) and (F) appear to present for it. So not only do we have a theory of what logic is: namely, the science that gives the general laws of consequence relations. We also have an account, entirely consonant with that theory, of what distinguishes broadly logical consequence from the other relations with which logic deals.

References

- Cartwright, R. 1997. “Singular Propositions.” *Canadian Journal of Philosophy, Supplementary Vol. 23*: 67–84.
- Copi, I.M. 1973. *Symbolic Logic*. 4th edition. New York: Macmillan.
- Davies, M. and L. Humberstone 1980. “Two Notions of Necessity.” *Philosophical Studies* 38: 1–30.
- Dretske, F. 1970. “Epistemic Operators.” *Journal of Philosophy* 67: 1007–1023.
- Dummett, M.A.E. 1981. *Frege: Philosophy of Language*. 2nd edition. London: Duckworth.
- Edgington, D. 2004. “Two Kinds of Possibility.” *Proceedings of the Aristotelian Society, Supplementary Vol. 78*: 1–22.
- Etchemendy, J. 1990. *The Concept of Logical Consequence*. Cambridge, Mass.: Harvard University Press.
- Evans, G. 1979. “Reference and Contingency.” *Monist* 62: 161–189.

²³ I make the comparison in Rumfitt 2010.

- Fine, K. 2002. "The Varieties of Necessity." In *Conceivability and Possibility*, edited by T. Szabó Gendler and J. Hawthorne, 253–282. Oxford: Clarendon Press.
- Frege, G. 1893. *Grundgesetze der Arithmetik*. Vol. 1. Jena: Pohle.
- Gentzen, G. 1935. "Untersuchungen über das logische Schliessen." *Mathematische Zeitschrift* 39: 176–210; 405–431. Translated as "Investigations into Logical Deduction" in *The Collected Papers of Gerhard Gentzen*, ed. and trans. by M.E. Szabó, 68–131. Amsterdam: North Holland.
- Gibbard, A. 2003. *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.
- Hacking, I. 1979. "What is Logic?" *Journal of Philosophy* 76: 285–319.
- Kneale, W. and M. Kneale. 1962. *The Development of Logic*. Oxford: Clarendon Press.
- Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.
- Lemmon, E.J. 1965. *Beginning Logic*. London: Nelson and Sons.
- Mates, B. 1965. *Elementary Logic*. New York: Oxford University Press.
- McFetridge, I. 1990. Logical Necessity: Some Issues. In *Logical Necessity and Other Essays*, edited by J. Haldane and R. Scruton, 135–154. London: Aristotelian Society.
- Mill, J.S. 1891. *A System of Logic, Ratiocinative and Inductive*. 8th ed. London: Longman.
- Moore, G.E. 1922. *Philosophical Studies*. London: Routledge and Kegan Paul.
- Peacocke, C. 1978. "Necessity and Truth Theories." *Journal of Philosophical Logic* 7: 473–500.
- Quine, W.V. 1951. "Two Dogmas of Empiricism." *Philosophical Review* 60: 20–43.
- . 1976. *The Ways of Paradox and Other Essays*. Revised and enlarged ed. Cambridge, Mass.: Harvard University Press.
- Restall, G. 2005. Multiple Conclusions. In *Logic, Methodology, and Philosophy of Science: Proceedings of the Twelfth International Congress*, 189–205. London: King's College Publications.
- Rumfitt, I. 2001. "Semantic Theory and Necessary Truth." *Synthese* 126: 283–324.
- . 2008. "Knowledge by Deduction." *Grazer Philosophische Studien* 77: 61–84.
- . 2010. Logical Necessity. In *Modality: Metaphysics, Logic and Epistemology*, edited by R.L.V. Hale and A. Hoffman, 35–64. Oxford: Clarendon Press.
- . forthcoming. "Objects of Thought." In *Meanings and Other Things: Essays in Honour of Stephen Schiffer*, edited by G. Ostertag. Cambridge, Mass.: MIT Press.
- Russell, B. 1919. *Introduction to Mathematical Philosophy*. London: Allen and Unwin.
- Shoemsmith, D.J. and T.J. Smiley 1978. *Multiple Conclusion Logic*. Cambridge: Cambridge University Press.
- Smiley, T.J. 1995. "A Tale of Two Tortoises." *Mind* 104: 725–736.

-
- Tarski, A. 1930. "Über einige fundamentale Begriffe der Metamathematik." *Comptes Rendues des séances de la Société des Sciences et des Lettres de Varsovie* Classe III 23: 22–29. Translated by J.H. Woodger as "On Some Fundamental Concepts of Metamathematics," in Tarski 1983.
- . 1936. "Über den Begriff der logischen Folgerung." *Actes du Congrès International de Philosophie Scientifique* 7: 1–11. Translated by J.H. Woodger as "On the Concept of Logical Consequence," in Tarski 1983.
- . 1983. *Logic, Semantics, Mathematics*. Edited by J. Corcoran. Indianapolis: Hackett.
- Tennant, N. 1997. *The Taming of the True*. Oxford: Clarendon Press.

TIMOTHY WILLIAMSON

Absolute Identity and Absolute Generality

The aim of this chapter is to tighten our grip on some issues about quantification by analogy with corresponding issues about identity on which our grip is tighter. We start with the issues about identity.

I

In conversations between native speakers, words such as ‘same’ and ‘identical’ do not usually cause much difficulty. We take it for granted that others use them with the same sense as we do. If it is unclear whether numerical or qualitative identity is intended, a brief gloss such as “one thing not two” for the former or “exactly alike” for the latter removes the unclarity. In this paper, numerical identity is intended. A particularly conscientious and logically aware speaker might explain what ‘identical’ means in her mouth by saying: “Everything is identical with itself. If something is identical with something, then whatever applies to the former also applies to the latter.” It seems perverse to continue doubting whether ‘identical’ in her mouth means *identical* (in our sense). Yet other interpretations are conceivable. For instance, she might have been speaking an odd idiolect in which ‘identical’ means *in love*, under the misapprehension that everything is in love with itself and with nothing else (narcissism as a universal theory).

Let us stick to interpretations on which she spoke truly. Let us also assume for the time being that we can interpret her use of the other words homophonically. We will make no assumption at this stage as to whether ‘everything’ and ‘something’ are restricted to a domain of contextually relevant objects. We can argue that ‘identical’ in her mouth is coextensive with ‘identical’ in ours. For suppose that an

object x is identical in her sense with an object y . By our interpretative hypotheses, if something is identical in her sense with something, then whatever applies to the former also applies to the latter. Thus whatever applies to x also applies to y . By the logic of identity in our sense (in particular, reflexivity), everything is identical in our sense with itself, so x is identical in our sense with x . Thus being such that x is identical in our sense with it applies to x . Consequently, being such that x is identical in our sense with it applies to y . Therefore, x is identical in our sense with y . Generalizing: whatever things are identical in her sense are identical in ours. Conversely, suppose that x is identical in our sense with y . By the logic of identity in our sense (in particular, Leibniz's Law), if something is identical in our sense with something, then whatever applies to the former also applies to the latter. Thus whatever applies to x also applies to y . By our interpretative hypotheses, everything is identical in her sense with itself, so x is identical in her sense with x . Thus being such that x is identical in her sense with it applies to x . Consequently, being such that x is identical in her sense with it applies to y . Therefore, x is identical in her sense with y . Generalizing: whatever things are identical in our sense are identical in hers. Conclusion: identity in her sense is coextensive with identity in our sense.¹

Of course, coextensiveness does not imply synonymy or even necessary coextensiveness. Thus we have not yet ruled out finer-grained differences in meaning between her use of 'identical' and ours. If we can interpret her explanation as consisting of logical truths, then, given that the principles that the argument invoked about identity in our sense (reflexivity and Leibniz's Law) are also logical truths, we can show that the universally quantified biconditional linking identity in her sense with identity in ours is a logical truth, so that coextensiveness is logically guaranteed.² If the relevant kind of logical truth is closed under the rule of necessitation from modal logic, then the necessitated universally quantified biconditional too is a logical truth, so that necessary coextensiveness is also logically guaranteed. But not even a logical

¹ The argument goes back to Quine (1961); see the reprinted version in Quine 1966 (178).

² The logic of indexicals is arguably not closed under the rule of necessitation (Kaplan 1989). Such problems do not seem to arise for (1) and (2).

guarantee of necessary coextensiveness suffices for synonymy: we have such a guarantee for ‘cat who licks all and only those cats who do not lick themselves’ and ‘mouse who is not a mouse,’ but they are not synonymous. Nevertheless, even simple coextensiveness excludes by far the worst forms of misunderstanding.

But now we must reconsider our homophonic interpretation of all our speaker’s other words. In her second claim “If something is identical with something, then whatever applies to the former also applies to the latter,” how much did ‘whatever’ cover? We can regiment her utterance as the schema (2) of first-order logic (for the record, we also formalize her first claim as (1)):

$$(1) \forall x xIx$$

$$(2) \forall x \forall y (xIy \rightarrow (A(x) \rightarrow A(y)))$$

Here ‘*I*’ symbolizes identity in her sense; $A(y)$ differs from $A(x)$ at most in having the free variable y in some or all places where $A(x)$ has the free variable x . Our speaker will instantiate (2) only by formulas $A(x)$ and $A(y)$ of her language. But our argument for coextensiveness in effect involved the inference from xIy and $x=x$ to $x=y$, where ‘=’ symbolizes identity in our sense. To use (2) for that purpose, we must take $A(x)$ and $A(y)$ to be $x=x$ and $x=y$ respectively. By what right did we treat $x=x$ and $x=y$ as formulas of her language, not merely of ours? Perhaps her language has no equivalent formulas, and both (1) and all instances of (2) in her language are true even though ‘*I*’ does not have the extension of identity in our sense.

We can be more precise. Let M be an ordinary model with a non-empty domain D for a first-order language L . Define a new model M^* for L as follows. The domain D^* of M^* contains all and only ordered pairs $\langle d, j \rangle$, where d is a member of D and j is a member of some fixed index set J of finite cardinality $|J|$ greater than one; pick a member $\#$ of J . If R is an n -adic atomic predicate of L , the extension of R in M^* contains the n -tuple $\langle \langle d_1, j_1 \rangle, \dots, \langle d_n, j_n \rangle \rangle$ if and only if the extension of R in M contains the n -tuple $\langle d_1, \dots, d_n \rangle$. If the constant c denotes d in M , c denotes $\langle d, \# \rangle$ in M^* . More generally, if the n -place function symbol f of L denotes a function φ in M , f denotes the function φ^* in M^* , where $\varphi^*(\langle d_1, j_1 \rangle, \dots, \langle d_n, j_n \rangle)$ is $\langle \varphi(d_1, \dots, d_n), \# \rangle$. Given this definition of M^* , it is routine to prove that exactly the same

formulas of L are true in M^* as in M .³ Now suppose that the dyadic atomic predicate I of L is interpreted by identity in M ; its extension there consists of all pairs $\langle d, d \rangle$, where d is in D . Thus (1) and all instances of (2) are true in M . Consequently, (1) and all instances of (2) are true in M^* . Nevertheless, I is not interpreted by identity in M^* , for the extension of I in M^* contains the pair $\langle \langle d, j \rangle, \langle d, k \rangle \rangle$ for any member d of D and members j and k of J . Thus in M^* everything has the relation for which I stands to $|J|$ things, and therefore to $|J|-1$ things distinct from itself, although one cannot express that fact in L . Nor does any formula, simple or complex, of L express identity in M^* , whether or not identity is expressed in M by some formula.⁴

³ Sketch of proof: Let a be any assignment of values in D^* to variables. Let $a^\wedge$ be the assignment of values in D to variables such that for any variable v , if $a(v)$ is $\langle d, j \rangle$ then $a^\wedge(v)$ is d . It is routine that for any term t , if t denotes $\langle d, j \rangle$ in M^* relative to a then t denotes d in M under $a^\wedge$ (by induction on the complexity of t). Thus any atomic formula $Rt_1, \dots, t_n$ is true in M^* under an assignment a if and only if it is true in M under $a^\wedge$, by definition of the extension of R in M^* . We show that any formula A of L is true in M^* under an assignment a if and only if it is true in M under $a^\wedge$ by induction on the complexity of A . The induction step for the truth-functores is trivial. For the universal quantifier, the induction hypothesis is this: for every assignment b of values in D^* , A is true in M^* under b if and only if it is true in M under $b^\wedge$. Suppose that $\forall x A$ is not true in M^* under an assignment a . Then, under some assignment b of values in D^* that differs from a at most over x , A is not true in M^* . By induction hypothesis, A is not true in M under $b^\wedge$. By construction, $b^\wedge$ differs from $a^\wedge$ at most over x . Thus $\forall x A$ is not true in M under $a^\wedge$. Conversely, suppose that $\forall x A$ is not true in M under $a^\wedge$. Then, under some assignment $a^!$ of values in D that differs from $a^\wedge$ at most over x , A is not true in M . Let b be the assignment of values in D^* like a except that $b(x)$ is $\langle a^!(x), \# \rangle$. Thus $b^\wedge$ is $a^!$, so A is not true in M under $b^\wedge$. By induction hypothesis, A is not true in M^* under b . Since b differs from a at most over x , $\forall x A$ is not true in M^* under a . Thus any formula A is true in M^* under an assignment a if and only if it is true in M under $a^\wedge$. Finally, we show that a formula is true in M^* if it is true in M . If A is not true in M^* , then, under some assignment a of values in D^* , A is not true in M^* , so A is not true in M under $a^\wedge$, so A is not true in M . Conversely, if A is not true in M , then, under some assignment $\$$ of values in D , A is not true in M ; but $\$$ is $a^\wedge$ for some assignment a of values in D^* (since we can set $a(v)$ to be $\langle \$ (v), \# \rangle$), so A is not true in M^* under a , so A is not true under a . The whole construction is adapted from the standard proof of the upward Löwenheim-Skolem theorem for first-order logic without identity. All it really requires is a homomorphism in a suitable sense from M^* onto M ; thus it is inessential that an equal and finite number of members of D^* are mapped to each member of D .

⁴ Proof: Suppose that $A(x,y)$ expresses identity in M^* : for every assignment a of values in D^* , $A(x,y)$ is true in M^* under a if and only if $a(x)$ is $a(y)$. For some member d of D , let a be an assignment of values in D^* such that both $a(x)$ and $a(y)$ are $\langle d, \# \rangle$. By hypothesis, $A(x,y)$ is true in M^* under a . Let b be an

We cannot rule out M^* by adding a sentence of L that is false in M^* to our theory of identity ((1) and (2)), for any such sentence will also be false in M , the “intended” model.

We may object that M^* is not a model for a first-order language with identity, precisely because it does not interpret any atomic predicate of the language by identity. What distinguishes first-order logic with identity from first-order logic without identity is that the former treats an atomic identity predicate as a logical constant. In standard first-order logic with identity, logical consequence is defined as truth-preservation in all models, and all models are stipulated to interpret that predicate by identity. Unintended interpretations of some basic mathematical terms can be excluded in first-order logic with identity but not in first-order logic without identity.

An example is the concept of linear (total) ordering. In first-order logic with identity, we standardly axiomatize the theory of (reflexive) linear orders (such as $\leq$ on the real numbers) by (3), (4) and (5):

- | | | |
|-----|--|-----------------|
| (3) | $\forall x \forall y \forall z ((xRy \ \& \ yRz) \rightarrow xRz)$ | (transitivity) |
| (4) | $\forall x \forall y (xRy \vee yRx)$ | (connectedness) |
| (5) | $\forall x \forall y ((xRy \ \& \ yRx) \rightarrow x=y)$ | (anti-symmetry) |

The models of this little theory are exactly those in which R is interpreted by a reflexive linear order (over the relevant domain). The use of the identity predicate in the anti-symmetry axiom (5) is essential. For if R is interpreted by a reflexive linear order, then the open formula $Rxy \ \& \ Ryx$ must express identity (over the relevant domain). But we have seen that in first-order logic without identity any theory with a model M has a model M^* as above in which no formula expresses identity, therefore in which R is not interpreted by a reflexive linear order. Consequently, no theory in first-order logic without identity has as models exactly those in which R is interpreted by a reflexive linear order. Since the models of first-order logic with and without identity

assignment of values like a except that $b(y)$ is $\langle d, \#\# \rangle$ for some member $\#\#$ of J distinct from $\#$. By hypothesis, $A(x,y)$ is not true in M^* under b , since $b(x)$ is not $b(y)$. By the argument of the previous footnote and in its notation: $A(x,y)$ is true in M^* under a if and only if it is true in M under $a^\wedge$; $A(x,y)$ is true in M^* under b if and only if it is true in M under $b^\wedge$. But $a^\wedge$ is $b^\wedge$. Thus $A(x,y)$ is true in M^* under a if and only if it is true in M^* under b . Contradiction.

differ only over the interpretation of the identity predicate, even in first-order logic with identity no theory axiomatized purely by sentences without the identity predicate has as models exactly those in which R is interpreted by a reflexive linear order.

The position is substantially the same for irreflexive linear orders. To axiomatize the theory of irreflexive linear orders (such as $<$ on the real numbers), we standardly retain axiom (3) but replace (4) and (5) by (6) and (7):

- | | |
|--|-------------|
| (6) $\forall x \forall y (xRy \vee yRx \vee x=y)$ | (linearity) |
| (7) $\forall x \forall y (xRy \rightarrow \sim yRx)$ | (asymmetry) |

The use of the identity predicate in the linearity axiom (6) is essential. For if R is interpreted by an irreflexive linear order, then the open formula $\sim Rxy$ & $\sim Ryx$ must express identity (over the relevant domain). In first-order logic without identity, any theory with a model has a model in which no formula expresses identity, therefore in which R is not interpreted by an irreflexive linear order. Consequently, no theory in first-order logic without identity has as models exactly those in which R is interpreted by an irreflexive linear order. Even in first-order logic with identity, no theory axiomatized purely by sentences without the identity predicate has as models exactly those in which R is interpreted by an irreflexive linear order.

First-order logic with identity is superior in expressive power to first-order logic without identity in mathematically central ways.⁵

Nevertheless, the appeal to first-order logic with identity may not resolve the doubts of those who take the problem of interpretation seriously. Indeed, it may strike them as cheating. For how do we know that the speaker whom we are trying to interpret is using a first-order language with identity at all? For example, how do we know that she is trying to talk about linear ordering? To pose the problem in less epistemic terms: what makes it the case that the speaker is using a first-order language with identity?

Quine has a short way with bloated models such as M^* . He excludes them by his methodology of interpretation, which requires us

⁵ For the view that identity is not a logical constant, which puts first-order logic with identity in an anomalous position, see Peacocke 1976.

to interpret the language in such a way that the strongest indiscernibility relation expressible in it is identity. Roughly speaking, he applies *a priori* the inverse of the operation that took M to M^* . As he says, we thereby “impose a certain identification of indiscernibles,” adding “but only in a mild way” (1960, 230). The “mildness” consists in this: indiscernibility in the relevant sense is the negation of weak discernibility, not of strong discernibility. Objects d and d^* in the domain of the interpretation are strongly discernible if and only if, for some open formula $A(x)$ of L with one free variable (x) and assignments a and a^* to variables of values in the domain, a^* is like a except that $a(x)$ is d while $a^*(x)$ is d^* , and the truth-value of $A(x)$ under a differs from its truth-value under a^* . The definition of weak discernibility is the same except that variables other than x are allowed to occur free in $A(x)$. For example, consider a language with just one atomic predicate, a dyadic one I , without constants or function symbols, and an interpretation with an infinite domain, over which I is interpreted by identity. Let d and d^* be distinct members of the domain. Then d and d^* are not strongly discernible, but they are weakly discernible by the formula xIy and assignments a and a^* , where $a(x)$ is d , $a(y)$ is d^* and a^* is like a except that $a(x)$ is d^* . In this case, Quine’s methodology does not erase distinctions between members of the domain: doing so here would involve collapsing the domain to a single object and so switching the formula $\forall x\forall y (xIy \rightarrow xIy)$ from false to true. The identification of indiscernibles as he apparently intends it has the hermeneutically appealing feature that it does not alter the truth-value of any formula.⁶

In other examples, Quine’s methodology has more radical effects on the model. For instance, consider another language with just two atomic predicates, the monadic F and G , without constants or function symbols, and an interpretation on which 1,000 members of the

⁶ The text does not follow Quine in inessential details. He speaks of the satisfaction of a formula with a given number of free variables by that number of objects in a given order, rather than of the truth of a formula under an assignment of objects to all variables. Quine 1960 (230) incorrectly claims that the relevant kind of discernibility is relative discernibility: satisfaction of an open formula with two free variables by the objects in only one order. Quine 1976 implicitly corrects the mistake. In the example in the text (taken from that article), no two objects are even relatively discernible. The text attributes a domain to the interpretation, perhaps contrary to Quine’s intentions, in order to make it clear that the argument here does not rely on contentious premises about unrestricted quantification.

domain are in the intersection of the extensions of F and G , just one is in the extension of F but not of G , 1,000,000 are in the extension of G but not of F , and just one in the extension of neither. Objects are weakly discernible if and only if either one is in the extension of F while the other is not or one is in the extension of G while the other is not. Thus Quine's "mild" identification of indiscernibles collapses the 1,000 objects in the intersection of the extensions of F and G into a single object, and the 1,000,000 objects in the extension of G but not F into another single object.⁷

Moreover, Quine's methodology does not preserve the truth-values of all formulas once we add generalized quantifiers to the language. For instance, let us add a binary quantifier M for "most" to the language in the last example, where $Mx (A(x); B(x))$ is true under an assignment a if and only if most (more than half) of the members d of the domain such that $A(x)$ is true under $a[d/x]$ are such that $B(x)$ is true under $a[d/x]$, where the assignment $a[d/x]$ is like a except that $a[d/x](x)$ is d . The addition of M to the language makes no difference to weak discernibility. On the original interpretation, the sentence $Mx (Fx; Gx)$ is true, because 1,000 of the 1,001 objects in the extension of F are in the extension of G , while $Mx (Gx; Fx)$ is false, because only 1,000 of the 1,001,000 objects in the extension of G are in the extension of F . By contrast, after the identification of indiscernibles, both sentences are false, because exactly one of the two objects in the extension of F is in the extension of G and exactly one of the two objects in the extension of G is in the extension of F . Moreover, no attempt to reinterpret 'M' as a logical quantifier other than "most" in line with the identification of indiscernibles would preserve the truth-values of all formulas. For the collapsed model is symmetrical between F and G : each applies to exactly one thing to which the other does not. Thus on any interpretation of 'M' as a logical quantifier $Mx (Fx; Gx)$ and $Mx (Gx; Fx)$ will receive the same truth-value in the collapsed model, whereas they have different truth-values in the new model (see Westerstahl 1989 for logical quantifiers).

Thus the ability of Quine's identification of indiscernibles to preserve the truth-values of all formulas depends on an unwarranted

⁷ See Wiggins 2001, 185 for discussion of a related example, attributed to Wallace (1964).

restriction of the language to the usual quantifiers $\forall$ and $\exists$.⁸ In the presence of other quantifiers, his identification of indiscernibles does not preserve truth-values, and so is hermeneutically unappealing. Of course, the example was a toy one; the expressive resources of the language were radically impoverished by comparison with any natural human language. Nevertheless, it shows that Quine's methodology does not provide an adequate solution to the problem of interpreting an identity predicate. For if independent considerations have not eliminated interpretations of a natural language on which its domain contains distinct indiscernibles, the use of Quine's methodology to do so risks imposing on the language an interpretation far less charitable to its speakers than some of the "bloated" interpretations are.

An alternative, anti-Quinean, proposal is to move from first-order to second-order logic. We could then replace the first-order schema (2) with a single second-order axiom:

$$(2+) \quad \forall x \forall y (xIy \rightarrow \forall P (Px \rightarrow Py))$$

In the usual models for higher-order logic, the second-order quantifier $\forall P$ is required to range in effect over all subsets of the first-order domain. Given any two objects d and d^* , some subset of the domain (for example, $\{d\}$) contains d but not d^* . Similarly, if one interprets $\forall P$ as a plural quantifier, there are some objects of which d is one and of which d^* is not one. Thus (2+), conjoined with (1), forces the predicate I to have the extension of identity over the domain. For practical purposes, one could even use the open formula $\forall P (Px \rightarrow Py)$ simply to *define* identity, although it is unlikely that second-order quantification is conceptually more basic than identity in any deep sense.⁹ But the appeal to second-order quantification may not satisfy those who are seriously

⁸ The addition of generalized quantifiers impacts on the "bloated" model M^* . No problem arises for "many," since the construction preserves the ratios between the cardinalities of the extensions of predicates (for the index set J was stipulated to be finite), but numerical quantifiers must be reinterpreted in order to give all formulas the same truth-values as in M : 'at least m ' is interpreted as "at least $m|J|$."

⁹ See also Shapiro 1991, 63. Since the values over which the second-order quantifier ranges are closed under complementation relative to the individual domain, strengthening the conditional in the *definiens* to a biconditional would make no difference.

worried about the problem of interpreting the identity predicate. For how do we know, or what makes it the case, that the second-order quantifier $\forall P$ should be interpreted in the standard way? Consider an interpretation of the first-order fragment of the language (with I as an atomic predicate) on which not all pairs of distinct members of the domain are weakly discernible, and the extension of I contains exactly those pairs of members of the domain that are not weakly discernible. We can now construct a non-standard interpretation of the full second-order language by stipulating that the range of the second-order quantifiers is to be restricted to those subsets of the domain with the property that if d is a member and d^* is not weakly discernible from d with respect to the first-order fragment then d^* is also a member (a similar stipulation is available for the plural interpretation).¹⁰ It can then be shown that, on this interpretation, objects are weakly discernible with respect to the full second-order language if and only if they are weakly discernible with respect to the first-order fragment. Consequently, not all distinct pairs of members of the first-order domain are weakly discernible with respect to the full second-order language. Nevertheless, (1) and (2+) come out true on such an interpretation even though I is not interpreted by identity over the domain. Thus invoking second-order logic only pushes the problem back to that of interpreting higher-order languages.¹¹

If we conceive the hermeneutic problem as purely epistemic—how do we know whether another means *identical*?—then we may suppose that it does not arise in the first-person case: how can I be mistaken in thinking that by ‘identical’ I mean *identical*? But if the problem is

¹⁰ The construction can be generalized to polyadic predicate variables and to orders greater than two, if desired.

¹¹ Such non-standard models are models in the sense of the non-standard semantics with respect to which Henkin (1950) proves the completeness of higher-order logic. As Shapiro (1991, 76) remarks, “Henkin semantics and first-order semantics are pretty much the same.” It is therefore no surprise that, once Henkin models are allowed, higher-order logic is no advance on first-order logic in solving the interpretation problem. Similarly, the substitution of Henkin semantics for the standard semantics throws away the advantages of second-order logic over first-order logic as a setting for mathematical theories. For example, the result that all models of second-order arithmetic are mutually isomorphic holds only for the standard semantics; non-standard models of first-order arithmetic can be simulated by appropriate Henkin models of second-order arithmetic.

constitutive—in virtue of what does another mean *identical* by ‘identical’?—then it presumably arises just as much in the first-person case: in virtue of what do I mean *identical* by ‘identical’? Despairing of an answer, someone might doubt the very conception of identity that underlies the question. One might even become a relativist about identity in the manner of Peter Geach, with a conception of a predicate’s playing the role of *I* relative to a given language, by verifying (1) and all instances of (2) in that language, but reject any conception of its playing the role absolutely, by verifying (1) and all instances of (2) in all possible extensions of the language.¹²

Such a reaction would be grossly premature, resting on no properly worked out, plausible account of interpretation. Questions of the form “In virtue of what do we mean *X* by ‘*X*’?” are notoriously hard to answer satisfactorily, no matter what is substituted for ‘*X*’ (Kripke 1982). It is therefore methodologically misguided to treat a particular expression (for instance, ‘identical’) as problematic merely on the grounds that the question is hard to answer satisfactorily for it.¹³ Of course, the details of the alternative interpretations and surrounding arguments depend on the nature of the expression at issue, but that should not cause us to overlook the generality of the underlying problem. It is extremely doubtful that the skeptical reaction yields anything coherent when generalized. In the particular case of identity, few have found Geach’s arguments for his local skepticism convincing or his relativism plausible. In any case, let us suppose that we do use ‘identical’ in an absolute way, and ask in virtue of what we do so. Given what has just been said, we should not expect more than a sketchy answer.

Somehow or other, ‘identical’ means what it does because we use it in the way that we do. A central part of that use concerns our inferential practice with the term, as summarized by (1) and (2) or (2+). What is crucial in our use of the first-order schema (2) (or a corresponding first-order inference rule) is that we do not treat it as exhausted by its instances in our current language. Rather, we have a

¹² Geach gives his views on identity in his 1967, 1972, 1980, 1991 and elsewhere. For critical discussion of them see Dummett 1981, 547–583 and 1993, Noonan 1997 and Hawthorne 2003, 111–123. See also Wiggins 2001, 21–54.

¹³ Geach can allow that in a suitable context ‘identical’ may mean *same F*, understood as identity relative to the informative sortal expression *F*; the question concerns the use of ‘identical’ in contexts that supply no such sortal.

general disposition to accept instances of (2) in extensions of our current language. That is not to say that in all circumstances in which we are or could be presented with an instance of (2) in an extension of our current language, we accept it. Obviously, we may reject it as a result of computational error, or die of shock at the sight of it, or roll our eyes as a protest against pedantry; some instances may be too long or complex to be presented to us at all. But the existence of a large gap of that kind between the disposition and the conditionals is the normal case for dispositions, including ordinary physical dispositions such as fragility and toxicity: external factors of all sorts (such as antidotes) can intervene between them and their manifestations. The link between the disposition to D if C and conditionals of the form "If C, it Ds" is of a much looser sort. The failure of some of the associated conditionals does not show the absence of the disposition.¹⁴ We have the general disposition because we respond, when we do, to the general form of the schema (2) (or of a corresponding inference rule) rather than treating each of its instances as an independent problem. Our non-intentionally described behavior alone does not make it intelligible why we count as responding to the actual form of (2) and not to some gerrymandered variant on it with *ad hoc* restrictions for cases beyond our ken: it is also relevant that the actual form is more natural than the gerrymandered one in a way that fits it for being meant (it is a "reference magnet").¹⁵

Our understanding of (2) as transcending the bounds of our current language is already suggested by our use of the phrase 'Leibniz's Law' for a single principle. We do not usually think of the phrase as ambiguously denoting lots of different principles, one for each language.

In the case of the second-order axiom (2+), what is crucial is that we do not treat the rule of universal instantiation for the second-order

¹⁴ On the relation between dispositions and conditionals see Martin 1994, Lewis 1997, Bird 1998 and Mumford 1998. For the specific application to rule-following see Martin and Heil 1998.

¹⁵ On the role of naturalness in the constitution of meaning see Lewis 1983. Of course, we may have to check putative instances of (2) to ensure that they really have the form that they appear to have and do not involve intensional or quotational contexts, shifts of reference in indexicals and so on. But such problems are not at the heart of the dispute between absolutists and relativists about identity.

quantifier as exhausted by its instances in our current language. Rather, we have a general disposition to accept instances of universal instantiation for the second-order quantifier in extensions of our current language. Again, the presence of the disposition is consistent with the failure of some of the associated conditionals. We have the general disposition because we respond to the general form of universal instantiation rather than treating each of its instances as a separate problem.

The sort of open-ended commitment just described is typical of our commitment to rules of inference. For example, my commitment to reasoning by disjunctive syllogism is not exhausted by my commitment to its instances in my current language; when I learn a new word, I am not faced with an open question concerning whether to apply disjunctive syllogism to sentences in which it occurs. Indeed, open-ended commitment may well be the default sort of commitment: one's commitment is open-ended unless one does something special to restrict it.

Open-ended commitment is just what is needed to reinstate the argument given at the beginning for a homophonic interpretation of another's use of the word 'identical,' given her commitment to (in effect) (1) and (2). We reason as follows. Let agents S and S^* speak distinct first-order languages L and L^* respectively. For simplicity, assume that S and S^* coincide in their logical vocabulary, with the possible exception of an identity predicate. Their interpretation of the other logical vocabulary is assumed to be standard; although the present style of argument can be extended to the other logical vocabulary, that is not our present concern. Let I be a predicate of L but not of L^* and I^* a predicate of L^* but not of L . Suppose that S has an open-ended commitment to (1) and (2), while S^* has an open-ended commitment to (1 *) and (2 *), the results of substituting I^* for I in (1) and (2):

- (1 *) $\forall x \ x I^* x$
- (2 *) $\forall x \forall y \ (x I^* y \rightarrow (A(x) \rightarrow A(y)))$

Now merge L and L^* into a single first-order language $L+L^*$ whose primitive vocabulary is the union of the primitive vocabularies of L and L^* . Thus we can treat (1) and (1 *) as sentences of $L+L^*$ and (2) and (2 *) as schemas of $L+L^*$. The interpretation of the logical vocabulary of $L+L^*$ is assumed to be standard, like that of L and L^* , again with the possible exception of an identity predicate. The quantifiers of

$L+L^*$ are interpreted as ranging over the intersection of the domains of the quantifiers of L and of L^* , for our present question is in effect whether I and I^* can diverge for objects over which both are defined. By the open-endedness of their commitments, S is committed to (1) as a sentence of $L+L^*$ and to all instances of (2) in $L+L^*$, while S^* is committed to (1 *) as a sentence of $L+L^*$ and to all instances of (2 *) in $L+L^*$. Here are instances of (2) and (2 *) respectively in $L+L^*$:

- (8) $\forall x\forall y (xIy \rightarrow (xI^*x \rightarrow xI^*y))$
 (8 *) $\forall x\forall y (xI^*y \rightarrow (xIx \rightarrow xIy))$

Reasoning in $L+L^*$, we deduce (9) from (1 *) and (8) and (9 *) from (1) and (8 *):

- (9) $\forall x\forall y (xIy \rightarrow xI^*y)$
 (9 *) $\forall x\forall y (xI^*y \rightarrow xIy)$

Thus, given the pooled commitments of S and S^* , I and I^* are coextensive over the common domain.

The result should not be interpreted as concerning only the extensions of I and I^* in a new context created by the fusion of L and L^* . For the open-ended commitments in play were incurred by S and S^* in using I and I^* in the original contexts for L and L^* respectively; that is the nature of open-endedness. Thus the result concerns the extensions of I and I^* in the original contexts for L and L^* too.

In the way just seen, (1) and the open-ended schema uniquely characterize identity (recall that the other logical vocabulary in (9) and (9 *) is being given its standard interpretation). A similar argument can be given for the second-order analogue of (2). The arguments are in fact a special case of a more general pattern of reasoning that shows all the usual logical constants to be uniquely characterized by the classical principles of logic for them.¹⁶

Of course, these remarks fall far short of a fully satisfying account of what makes 'identical' mean *identical*. The connection between which logical principles a speaker *accepts* for a given expression and

¹⁶ For further discussion and references see Harris 1982, Williamson 1987/88 and McGee 2000.

which logical principles are *correct* (true or truth-preserving) for that expression is quite loose; no reasonable principle of charity in interpretation guarantees freedom from logical error. Misguided philosophers who reject standard logical principles for 'identical' probably still mean *identical* by the word, because they continue to use it as a word of the common language.¹⁷ It is a fallacy to reason from the premise that someone has a wildly deviant theory to the conclusion that they speak a deviant language. Perhaps even more misguided philosophers could argue themselves into the view that the logic of the phrase 'in love' comprises analogues of the standard logical principles for 'identical,' while still meaning *in love* rather than *identical* by the phrase, because they continue to use it as a phrase of the common language. Nevertheless, such examples do not suggest that the account above of how 'identical' means *identical* does not at least point in the right direction, for they may still be parasitic on a loose underlying connection between inferential practice and meaning.

Naturally, the account will not help if it is incoherent, as Geach would claim it to be. According to him, generalizing over all predicates in all possible extensions of the language generates semantic paradoxes (1972, 240; 1991, 297). Now the foregoing account does not assume that speakers who reason with (1) and (2) or (2+) must themselves have a conception of all predicates in all possible extensions of the language. For the sentences involved in such reasoning need not be metalinguistic. If speakers already have metalinguistic vocabulary in the language, they can use it in what they substitute for $A(x)$ and $A(y)$ in (2), but that does not imply that the schema itself is distinctively metalinguistic. Agents may lack the conceptual apparatus necessary to give a reflective account of their own practices. However, the foregoing theoretical account does deploy something reminiscent of the conception of all predicates in all possible extensions of the language on its own behalf, in explaining the nature of speakers' open-ended commitments. Thus Geach's charge is at least relevant.

Unfortunately, Geach does not bother to argue in detail that the theorist of absolute identity really requires conceptual resources powerful enough to generate semantic paradoxes. In fact, when we consider identity over a given set domain, we need only generalize over

¹⁷ See Williamson 2003b for arguments of this type.

all subsets of that domain, or over all subsets of the Cartesian product of the domain with itself.¹⁸ For the unique characterization argument above, we need merely consider expansions of the language by a single dyadic atomic predicate I^* , whose extension is a subset of the Cartesian product of the original domain with itself. Similarly, schema (2) forces I to have the extension of identity over the domain as soon as we consider its instances in expansions of the language by a single monadic atomic predicate, whose extension is a subset of the original domain. In standard (Zermelo-Fraenkel) set theory, sets are closed under the power set operation and the formation of Cartesian products. Thus quantification over subsets of the original set domain or of its Cartesian product with itself is quantification over another set domain. Consequently, the kind of generalization required for the theorist of absolute identity over a given set domain is of a quite harmless sort. It poses no serious threat of semantic or set-theoretic paradox. Some mysteries about the power set operation remain unsolved, notably Cantor's continuum problem (how many subsets has the set of natural numbers?), but they are not paradoxes. In any case, they are largely independent of the application to identity, for they concern the size of the whole power set, whereas in order to characterize identity it suffices to have just the singleton sets of members of the original domain; there are no more singletons of members than members. Geach's argument for the incoherence of absolute identity theory does not withstand attention.

Suppose that the foregoing account of our grasp of absolute identity is correct, as far as it goes. What does it suggest about our grasp of absolute generality, generality over absolutely everything, without any explicit or implicit restrictions whatsoever?

II

Sympathetic readers will have felt little difficulty in understanding the words "absolute generality, generality over absolutely everything, without any explicit or implicit restrictions whatsoever": but in principle

¹⁸ The Cartesian product of sets X and Y is the set of all ordered pairs whose first member belongs to X and second member belongs to Y .

those words are open to alternative interpretations. ‘Absolute’ might be read as itself relative to a background contextually supplied standard, and the quantifiers ‘any’ and ‘whatsoever’ over restrictions as themselves contextually restricted. One can find oneself saying “By ‘everything’ I mean *everything*” with the same desperate intensity with which one may say “By ‘identical’ I mean *identical*.” How deep does the similarity of the interpretative challenges go?

Let us start with the question of unique characterization. Here are standard rules for a (first-order) universal quantifier:¹⁹

$\forall$ -Introduction Given a deduction of A from some premises, one may deduce $\forall v A(v/t)$ from the same premises, where $A(v/t)$ is the result of replacing all occurrences of the individual constant t in the formula A by the individual variable v , provided that no such occurrence of v is bound in $A(v/t)$ and that t occurs in none of the premises.

$\forall$ -Elimination From $\forall v A$ one may deduce $A(t/v)$, where $A(t/v)$ is the result of replacing all free occurrences of the individual variable v in the formula A by the individual constant t .

Consider a universal quantifier $\forall$ in a language L governed by those rules, and another universal quantifier $\forall^*$ in a language L^* governed by exactly parallel rules, $\forall^*$ -Introduction and $\forall^*$ -Elimination. Suppose that the commitment of speakers of L and L^* to their principles is open-ended in the way discussed above for the case of the identity rules. Merge L and L^* into a single language $L+L^*$, whose primitive vocabulary is the union of the primitive vocabularies of L and L^* . Let A be a formula of $L+L^*$ in which the individual constant t does not occur and no variable except v occurs free. We reason in $L+L^*$. From $\forall v A$ we can deduce $A(t/v)$ by $\forall$ -Elimination. Therefore, since t does not occur in the premise and A is the result of replacing all occurrences of t in $A(t/v)$ by v , and no such occurrence of v thereby becomes bound

¹⁹ For simplicity, functional terms are ignored. To qualify a variable or constant as “individual” is just to say that it occupies singular term position.

in A , from $\forall v A$ we can deduce $\forall^* v A$ by $\forall^*$ -Introduction. Conversely, from $\forall^* v A$ we can deduce $A(t/v)$ by $\forall^*$ -Elimination, and therefore $\forall v A$ by $\forall$ -Introduction. Thus, given the pooled commitments of speakers of L and L^* , the two quantifiers are logically equivalent.²⁰

The result should not be interpreted as concerning only the reference of $\forall$ and $\forall^*$ in a new context created by the fusion of L and L^* . For the open-ended commitments in play were incurred by S and S^* in using $\forall$ and $\forall^*$ in the original contexts for L and L^* respectively; that is the nature of open-endedness. Thus the result concerns the reference of $\forall$ and $\forall^*$ in the original contexts for L and L^* too.

Observe that the argument for unique characterization did not proceed by semantic analysis of the quantifier. It did not invoke the idea of unrestricted generality. In particular, the argument was not that only an unrestricted interpretation of $\forall$ validates $\forall$ -Introduction and $\forall$ -Elimination. Rather, it was a syntactic argument for interderivability. Thus it is not circular to use the unique characterization result to support the claim that we have an idea of unrestricted generality.

Nevertheless, it is tempting to suspect the argument for unique characterization of sophistry. For $\forall$ -Introduction and $\forall$ -Elimination are standard rules for a universal quantifier in standard first-order logic, for which the standard model theory interprets the quantifier as restricted to the domain of a model. It may therefore look as though the argument must prove too much, since $\forall$ -Introduction and $\forall$ -Elimination are valid if $\forall$ is interpreted over a domain D , while $\forall^*$ -Introduction and $\forall^*$ -Elimination are valid even if $\forall^*$ is interpreted over a distinct domain D^* .

In that form, the objection is unthreatening, for it neglects the stipulation that the commitment of speakers of L and of L^* to the respective quantifier rules is open-ended in the sense explained in part I. If $\forall$ is restricted to a domain D , then speakers of L do not have an open-ended commitment to $\forall$ -Elimination, even if the latter has no counter-instance in L , since it has the potential for a counter-instance with a new term t that denotes something outside D in a language such as $L+L^*$. In the setting of $L+L^*$, $\forall$ -Elimination would require an extra premise involving t to the effect, concerning what t denotes, that it

²⁰ See McGee 2000 and Rayo 2003 for related discussion. For opposed views see Dummett 1981 and Glanzberg 2004.

belongs to D . Then t would occur in one of the premises from which $A(t/v)$ was deduced, so the condition for the application of $\forall^*$ -Introduction would not be met. Of course, if $\forall^*$ were restricted too, to a domain D^* , then one might modify $\forall^*$ -Introduction accordingly, by allowing t to occur in one extra premise of the envisaged deduction of A to the effect, concerning what t denotes, that it belongs to D^* . With no guarantee that D includes D^* , however, the deduction of $\forall^*v A$ from $\forall v A$ could not be carried through. The converse deduction faces an exactly analogous problem. It may sometimes be hard to know whether a given speaker's commitment is open-ended, but the considerations of part I indicated that open-ended commitment to a rule is a genuine, recognizable phenomenon, and no reason has emerged to view the quantifier rules as exceptional in that respect. Indeed, as before, open-ended commitment may be the default sort of commitment to $\forall$ -Introduction and $\forall$ -Elimination; it would be implausible to suggest that all speakers are always doing something special to override the default.

The pertinent objection is not to the argument for unique characterization from the open-ended understanding of the quantifier rules. Rather, it is to the open-ended understanding of the quantifier rules itself. More precisely: it is not straightforward that the open-ended versions $\forall$ -Introduction and $\forall$ -Elimination are really valid on the unrestricted reading of the quantifier.

Let us take $\forall$ -Elimination first. Free logicians will object to it that, even if $\forall$ is supposed to be unrestricted, the rule is too strong because the individual constant t may be an empty name. For example, $\forall$ -Elimination enables us to derive from the logically true premise $\forall y \exists x x=y$ the conclusion $\exists x x=t$, which is false on the unrestricted reading of the quantifier if t denotes nothing whatsoever.

One response to that objection to $\forall$ -Elimination is that the only role of t in the unique characterization argument is as an arbitrary name, which functions like a free variable. The success of any empirical or conceptual process of reference-fixing for t is irrelevant to the argument. Thus, if one shares the free logicians' qualms, one can replace 'individual constant' by 'arbitrary name' or 'free variable' throughout $\forall$ -Introduction and $\forall$ -Elimination, and envisage t as denoting merely relative to an assignment.

A less concessive response can also be made. In assessing validity, our concern is with truth-preservation only when the relevant formulas

are fully interpreted. For a sentence to be fully interpreted, it is not enough that it is a meaningful formula of the language; it must also express a proposition as used in the relevant context. For example, although ‘This is that’ is a meaningful sentence of English, it fails to express a proposition in a context in which no reference has been assigned to the demonstratives ‘this’ and ‘that.’ It would be foolish to object to the usual introduction rule for disjunction (deduce a disjunction from any of its disjuncts) that it takes one from the true premise “ $2+2=4$ ” to the conclusion “ $2+2=4$ or this is that,” which is not true when no reference has been assigned to ‘this’ and ‘that’ because it expresses no proposition (plausibly, a disjunction expresses a proposition only if each of its disjuncts does). Truth-preservation is required only once any singular terms in the argument have been assigned a reference. Consequently, the free-logical objection to $\forall$ -Elimination fails. Let us adopt this conception of validity and therefore leave $\forall$ -Elimination unmodified.

“Inclusive” logicians object to $\forall$ -Elimination because it does not allow for the empty domain: given axiom (1) for identity, one can prove $\exists x x=x$ (as an abbreviation of $\sim\forall x \sim x=x$), which is false in that domain.²¹ But once the language contains unbound singular terms, as ours does, then it cannot be fully interpreted in the sense just sketched over the empty domain. For such languages, our notion of validity excludes the empty domain. More controversially, one can argue that $\exists x x=x$ is a logical truth on the unrestricted interpretation of the quantifiers, by appeal to Tarski’s model-theoretic account of logical truth: it is true on all models (interpretations) that preserve the intended interpretations of the logical constants in it, for it is true and it contains no nonlogical constituents.²² Since Tarski (1936) understood a model as an assignment of reference to the nonlogical atoms of the language (more exactly, as an assignment of values to variables, which replace those atoms), his treatment of interpretation is consistent with the

²¹ Logic for the empty domain is somewhat trickier than the remark in the text indicates; see Williamson 1999b.

²² See Williamson 1999a and Rayo and Williamson 2003 for this approach to the logic of unrestricted quantification. For reasons explained in the latter, the first-order quantification over interpretations in the text is a loose rendering of the higher-order quantification that is needed for an accurate metalogic of unrestricted quantification.

notion above of a fully interpreted formula. The relevance of Tarski's conception of logical truth and logical consequence to the logic of unrestricted quantification will be discussed more fully below.

Given an appropriate notion of validity, open-ended $\forall$ -Elimination is valid on the unrestricted reading of the quantifier. What of $\forall$ -Introduction? The obvious worry is this. Suppose that $\forall$ is unrestricted while $\forall^*$ is restricted to a domain D^* , and that the term t is constrained by a "meaning postulate" to be in D^* . Thus from $\forall^*x x \in D^*$ alone we can infer $t \in D^*$ by the restricted elimination rule for $\forall^*$; since t does not occur in the premise, $\forall$ -Introduction therefore permits us to infer $\forall x x \in D^*$, which is false because D^* does not contain absolutely everything, from the true $\forall^*x x \in D^*$. What has gone wrong here is that $t \in D^*$ was not *freely* deduced from $\forall^*x x \in D^*$, in a sense of 'free' that has nothing to do with free logic. The deduction was unfree in the sense that it invoked special rules that required t to satisfy constraints beyond simply being a singular term that occurs in none of the premises. That is still a syntactic feature of the deduction. Let us therefore read 'deduction' in $\forall$ -Introduction as "free deduction," and 'deduced' in both $\forall$ -Introduction and $\forall$ -Elimination as "freely deduced." With that understanding, both $\forall$ -Introduction and $\forall$ -Elimination are valid on the unrestricted reading of the quantifier.

Given that the two rules are materially valid on the unrestricted reading, someone might still worry that they are not strictly logically valid, because the unrestricted reading of the quantifier is not part of its logic. The discussion so far has been framed in terms of the background assumption that the unrestricted universal quantifier as such should be classified as a logical constant, subject to rules of inference that exploit its unrestrictedness. For those who admit the coherence of unrestricted quantification, the salient alternative is to have as a logical constant a universal quantifier such that, for any things whatsoever, for purposes of defining logical truth and logical consequence the quantifier can be interpreted as ranging over those things and nothing else (interpretations here play the role of models). It does not matter whether there are too many of those things to form a set or set-like domain. The quantifier can be legitimately interpreted as ranging over all things whatsoever, or over all sets whatsoever, but it can also be legitimately interpreted as ranging over just the books on my shelves. A conclusion is a logical consequence of some premises only if, however they

are legitimately interpreted, the conclusion is true if the premises are.²³ On this view, open-ended $\forall$ -Elimination is invalid, since in a language such as $L+L^*$ with different sorts of quantifier a term t may denote something in the domain of one of those other quantifiers that is not in the domain of $\forall$ on some unintended interpretation. Let the *constant account* be that on which the quantifier is mandatorily interpreted as unrestricted and the *variable account* be that on which, for any things, it can legitimately be interpreted as ranging over just those things. Both accounts are framed within a broadly Tarskian approach to the concept of logical consequence.

The difference between the two accounts has dramatic implications for first-order logic. Let $\exists n$ be the usual first-order formalization of the claim that there are at least n things, where $n > 1$. On the variable account, $\exists n$ is not a logical truth, because the quantifier can legitimately be interpreted as ranging over fewer than n things. On the constant account, $\exists n$ is a logical truth, because the quantifier must be interpreted as ranging over absolutely everything and there are in fact at least n things: for example, at least n symbols occur in $\exists n$ itself. The sentences $\exists n$ for all natural numbers n turn out to exhaust the extra logical consequences generated by the constant account, in the sense that the result of adding them as extra axioms to first-order logic can be proved complete as well as sound on the constant account.²⁴ By

²³ Cartwright (1994) takes this view.

²⁴ See Friedman 1999, Williamson 1999a, and Rayo and Williamson 2003. The result depends on a global choice assumption; Friedman discusses non-theorems of this logic that are logically true on some anti-choice assumptions. As an alternative to axioms of the form $\exists n$, the proponent of the constant account can use the following structural *Rule of Atomic Freedom*. Let $\Gamma \cup \{\gamma\}$ be a set of sentences, and $\Delta \cup \{\delta\}$ ($\delta \notin \Delta$) a set of atomic sentences each consisting of a non-logical predicate that does not occur in Γ and individual non-logical constants that do not occur in Γ such that $\Gamma, \Delta \vdash \delta$; then $\Gamma \vdash \gamma$ (Γ is inconsistent; it entails everything). To see why this rule preserves validity on the constant account, suppose that all members of Γ are true on some interpretation. Then all members of $\Gamma \cup \Delta \cup \{\sim\delta\}$ are also true on some interpretation, for we can stipulate that each constant in $\Delta \cup \{\sim\delta\}$ denotes itself and that the extension of each n -place predicate in $\Delta \cup \{\sim\delta\}$ is the set of n -tuples of singular terms with which it is concatenated in Δ ; thus every member of Δ is true, and δ is false because $\delta \notin \Delta$ (a simplified Henkin construction, which does not itself assume the existence of infinitely many things). Since the vocabulary of Γ is disjoint from that of $\Delta \cup \{\sim\delta\}$, it can be interpreted as originally; thus every member of Γ is true on the new interpretation too (this part of the argument

contrast, the variable account is logically conservative: it delivers exactly the same logical consequence relation for first-order logic as does the standard model theory with set domains.²⁵

The difference between the two accounts is robust. Even if we relativize interpretations to various parameters for the context of utterance or circumstance of evaluation, the sentences $\exists n$ still all come out as logical truths on the constant account, because the unrestricted reading of the quantifier forbids us to interpret it as ranging only over a domain associated with the context of utterance or circumstance. Someone might reply that if the domain contains everything that exists in the relevant possible world then the restriction is merely apparent, because in that world there is nothing else to quantify over. But that objection in effect treats the semantic clause for $\forall$ as though it were a misleading approximate translation of a more fundamental semantic clause in which an unrestricted universal quantifier of a more fundamental modal meta-language occurs within the scope of a modal operator. But the Tarskian framework for the theory of logical consequence is not a modal one. It defines logical consequence without using modal operators, interpreted metaphysically or epistemically. The non-modal meta-language should therefore be taken at face value: a semantic clause according to which $\forall$ ranges only over the domain of some world is inconsistent with the unrestricted reading, just as it appears to be. Indeed, it is part of Tarski's great achievement to have cleanly separated the concept of logical consequence from metaphysical and epistemic clutter. Not that there is anything wrong with metaphysical and epistemic modalities in their place: but it is methodologically wrong-headed to mix them up with the simple but powerful

would not work on the variable account, since there may be more constants in $\Delta \cup \{\sim\delta\}$ than things quantified over on the original interpretation). By contraposition, Atomic Freedom preserves validity. To see how to use Atomic Freedom to derive all sentences of the form $\exists n$, it suffices to look at the case $n=3$. Let R be a triadic predicate, a , b and c distinct constants, $\Gamma = \{\sim\exists 3\}$, $\Delta = \{Raac, Raba, Rabb\}$ and $\delta = Rabc$. By ordinary first-order reasoning, $\sim\exists 3, Raac, Raba, Rabb \vdash Rabc$ (unless there are at least three things, the three constants cannot all have distinct denotations); therefore, by Atomic Freedom, $\sim\exists 3 \vdash \perp$, so $\vdash \exists 3$. Note that the rule of Atomic Freedom is formulated without reference to any particular logical constant (contrast axioms of the form $\exists n$). Another way to think of the constant account is therefore as freeing up the interpretation of atomic formulas.

²⁵ See Cartwright 1994; the argument goes back to Kreisel 1967.

non-modal concept of logical consequence that Tarski painstakingly isolated, which compels and rewards investigation in its own right. This chapter works within the Tarskian paradigm.

The dispute between the variable and constant accounts raises deep questions about the metaphysical and epistemological status of logic that we cannot hope to answer here. But the analogy with identity does help us to see what is wrong with one argument against the constant account. It is sometimes urged that the variable account is preferable because it has greater generality, since every legitimate interpretation on the constant account is also legitimate on the variable account (for it allows us to interpret the quantifier as ranging over absolutely everything), but not *vice versa*. However, there is an analogous argument against first-order logic with identity, according to which first-order logic without identity is preferable because it has greater generality, since every legitimate interpretation in first-order logic with identity is also legitimate in first-order logic without identity (for it allows us to interpret a dyadic predicate by identity), but not *vice versa*. The latter argument clearly fails, because the point of making identity a logical constant is to capture its distinctive logic by excluding unintended interpretations. No significant generality is thereby lost, because all the other interpretations can be shifted to other dyadic predicates. The former argument against the constant account fails similarly, because the point of making the unrestricted universal quantifier a logical constant is to capture its distinctive logic by excluding unintended interpretations. No significant generality is thereby lost, because all the other interpretations can be captured by complex restricted quantifiers consisting of the simple unrestricted quantifier and a restricting predicate.

Of course, we have some sense of which expressions deserve to be treated as logical constants: very roughly, those whose meaning is “purely structural.”²⁶ By that standard, the unrestricted universal quantifier is at least as good a candidate as identity is. Moreover, like identity, the unrestricted quantifier has the kind of stark simplicity in meaning that we seek in a logical constant that is to be treated as basic (some purely structural meanings are very complicated). Although unrestricted quantification is less central to mathematical reasoning than

²⁶ Tarski (1986) proposes the more precise criterion of invariance under all permutations of individuals in this spirit.

identity is, it does enable us to capture the generality that principles of a set theory with ur-elements (non-sets) such as ZFU need if mathematics is to have its full range of applications: for absolutely *any* objects x and y , there is a set of which x and y are members, for example.

In using absolute identity to support absolute generality, we must be careful to check that the latter does not squash the former. For Geach's arguments against the coherence of absolute identity look superficially more formidable in the context of absolute generality. The previous section considered absolute identity over a set-sized domain; despite Geach's threats, no danger of paradox arises in characterizing it by quantifying over subsets of the domain, or plurally over members of the domain, or the like.²⁷ But if our first-order quantifiers are absolutely unrestricted, then the interpretation of the corresponding second-order quantifiers is a much trickier business. There is no set domain over whose subsets they could range. Indeed, any attempt to interpret them in terms of values of the second-order variables will generate a version of Russell's paradox, given an adequately strong comprehension principle concerning the existence of such values, since an absolutely unrestricted first-order quantifier must range over them too. But not even considerations of this kind can rescue the charge of incoherence against absolute identity. First, the paradox results from the attempt to interpret the second-order quantifiers of the object-language in a first-order meta-language, by first-order quantification over sets. If one interprets the second-order quantifiers more faithfully, in a second-order meta-language, by second-order quantification read plurally or in some other non-first-order way, then no paradox results. Second, even if one interprets the second-order quantifiers as ranging over "small" sets, with an appropriately qualified comprehension principle, that suffices for characterizing identity, although not for all other purposes. Third, if absolute identity is coherent for each set-sized domain, then it is simply coherent: for any objects o and o^* , absolute identity is coherent over the set-sized domain $\{o, o^*\}$ by hypothesis, which is all that we need coherently to ask whether o and o^* are absolutely identical.

²⁷ For the plural interpretation see several of the essays in Boolos 1998 (and 48–49 and 54 for brief remarks on the logic of identity). Williamson 2003a argues in favor of an interpretation that takes more seriously the idea of quantification into predicate position.

If there is a threat of paradox, it comes from the idea of absolute generality, not from that of absolute identity. A “paradox” here is a proof of an explicit contradiction from premises to which generality absolutists are committed by rules of inference to which they are also committed, something like the Russell or Burali-Forti paradox. The greatest strength of generality relativism is the suspicion that generality absolutism is ultimately inconsistent because it leads to such a paradox.

Generality relativists also tend to use a second sort of argument: that generality absolutism is inarticulate, in the sense that whatever utterances generality absolutists assent to or dissent from in trying to articulate their position, on some generality relativist interpretations all the assents were to truths and all the dissents from falsehoods (according to the generality relativist). In that sense, absolutism about identity is also inarticulate: whatever utterances identity absolutists assent to or dissent from in trying to articulate their position, on some identity relativist interpretations all the assents were to truths and all the dissents from falsehoods (according to the identity relativist). It does not follow that absolute identity is inexpressible, for all those interpretations were incorrect: as explained in section I, they misidentified our dispositions to make inferences using the identity predicate. Similarly, if generality absolutism is inarticulate, it does not follow that absolute generality is inexpressible, for the generality absolutist can argue that all the generality relativist interpretations were incorrect, because they misidentified our dispositions to make inferences using the universal quantifier. The generality absolutist may endorse a reflection principle to the effect that any quantified sentence true on the intended unrestricted interpretation of the quantifier is also true on some unintended restricted interpretation of the quantifier. Nevertheless, such a result does make the generality absolutism and its negation somewhat elusive for purposes of theoretical dispute.

It may be less widely appreciated than it should be that the generality relativist cannot combine the two objections to generality absolutism, by charging that it is *both* inconsistent *and* inarticulate. For suppose that generality absolutism is both inconsistent and inarticulate. Since it is inconsistent, there is a proof of an explicit contradiction from premises to which generality absolutists are committed by rules of inference to which they are committed. By hypothesis, generality

absolutism is also inarticulate, so on some generality relativist interpretation the premises of the proof are true (according to the generality relativist) and the rules of inference are truth-preserving (according to the generality relativist). Thus the generality relativist is committed to the truth of the conclusion of the proof on the generality relativist interpretation. But the conclusion is a contradiction, and so is not true even on that interpretation (according to the generality relativism). Thus generality relativism is inconsistent. To sum up: if generality absolutism is inarticulate, then it is inconsistent only if generality relativism is also inconsistent. Therefore, the generality relativist is ill-advised to accuse generality absolutism of being both inconsistent and inarticulate. In effect, the assumption that generality absolutism is inarticulate yields a consistency proof for generality absolutism relative to generality relativism.

Perhaps we can generalize the result. Given any argument against generality absolutism, why cannot it be reinterpreted as an argument against something that the generality relativist accepts, if generality absolutism really is inarticulate?

Generality relativists seem to be faced with a choice. They can drop the charge of inarticulacy, try to explain why it seemed compelling, and then, treating generality absolutism as an articulated theory, try to prove a contradiction in it. If they succeed in that, they win (dialetheism is a fate worse than death). But if they cannot produce such a proof, then they had better drop the charge of inconsistency: put up or shut up. Alternatively, they can drop the charge of inconsistency right away, and press the charge of inarticulacy. But that charge is hardly damaging to absolutism about generality, for it applies equally to absolutism about identity.²⁸

²⁸ This paper has been previously published in A. Rayo and G. Uzquiano, eds., *Absolute Generality* (Oxford: Clarendon Press, 2006), 369–389. Thanks to Kit Fine, Øystein Linnebo, Agustín Rayo and Gabriel Uzquiano for helpful written comments on a draft of this chapter. A version of the material was presented at the conference at the Central European University in Budapest; thanks also to the audience there, especially the commentator Katalin Farkas, for useful discussion.

References

- Bird, A. 1998. "Dispositions and Antidotes." *Philosophical Quarterly* 48: 227–234.
- Boolos, G. 1998. *Logic, Logic, and Logic*. Cambridge, Mass.: Harvard University Press.
- Cartwright, R. 1994. "Speaking of Everything." *Noûs* 28: 1–20.
- Dummett, M.A.E. 1981. *Frege: Philosophy of Language*. 2nd ed. London: Duckworth.
- . 1993. "Does Quantification Involve Identity?" In *The Seas of Language*. Oxford: Clarendon Press.
- Friedman, H. 1999. "A Complete Theory of Everything: Validity in the Universal Domain." Lecture note. <http://www.math.osu.edu/~friedman/pdf/ACompThyEver101099.pdf> (accessed February 9, 2011)
- Geach, P.T. 1967. "Identity." *Review of Metaphysics* 21: 3–12.
- . 1972. *Logic Matters*. Oxford: Blackwell.
- . 1980. *Reference and Generality*. 3rd ed. Ithaca, N.Y.: Cornell University Press.
- . 1991. "Replies." In *Peter Geach: Philosophical Encounters*, edited by H.A. Lewis. Dordrecht: Kluwer.
- Glanzberg, M. 2004. "Quantification and Realism." *Philosophy and Phenomenological Research* 69: 541–572.
- Harris, J.H. 1982. "What's So Logical about the 'Logical' Axioms?" *Studia Logica* 41: 159–171.
- Hawthorne, J. 2003. "Identity." In *The Oxford Handbook of Metaphysics*, edited by M. Loux and D. Zimmerman. New York: Oxford University Press.
- Henkin, L. 1950. "Completeness in the Theory of Types." *Journal of Symbolic Logic* 15: 81–91.
- Kaplan, D. 1989. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and Other Indexicals." In *Themes from Kaplan*, edited by J. Almog, J. Perry and H. Wettstein. New York: Oxford University Press.
- Kreisel, G. 1967. "Informal Rigour and Completeness Proofs." In *Problems in the Philosophy of Mathematics*, edited by I. Lakatos. Amsterdam: North-Holland.
- Kripke, S. 1982. *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.
- Lewis, D. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61: 343–377.
- . 1997. "Finkish Dispositions." *Philosophical Quarterly* 47: 143–158.
- Martin, C.B. 1994. "Dispositions and Conditionals." *Philosophical Quarterly* 44: 1–8.
- Martin, C.B. and J. Heil. 1998. "Rules and Powers." *Philosophical Perspectives* 12: 283–312.
- McGee, V. 2000. "Everything." In *Between Logic and Intuition*, edited by G. Sher and R. Tieszen. Cambridge: Cambridge University Press.

- Mumford, S. 1998. *Dispositions*. New York: Oxford University Press.
- Noonan, H. 1997. "Relative Identity." In *A Companion to the Philosophy of Language*, edited by B. Hale and C. Wright. Oxford: Blackwell.
- Peacocke, C. 1976. "What is a Logical Constant?" *Journal of Philosophy* 73: 221–240.
- Quine, W.V. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.
- . 1961. "Reply to Professor Marcus." In "Modalities and Intensional Languages." *Synthese* 13: 323–330.
- . 1966. *The Ways of Paradox and Other Essays*. New York: Random House.
- . 1976. "Grades of Discriminability." *Journal of Philosophy* 73: 113–116.
- Rayo, A. 2003. "When Does 'Everything' Mean *Everything*?" *Analysis* 63: 100–106.
- Rayo, A. and T. Williamson. 2003. "A Completeness Theorem for Unrestricted First-Order Languages." In *Liars and Heaps*, edited by J.C Beall. Oxford: Clarendon Press.
- Shapiro, S. 1991. *Foundations without Foundationalism: A Case for Second-order Logic*. Oxford: Clarendon Press.
- Tarski, A. 1936. "O pojęciu wynikania logicznego." *Przegląd Filozoficzny* 39: 58–68. Translated by J.H. Woodger as "On the Concept of Logical Consequence," in *Logic, Semantics, Metamathematics*, 2nd ed. Indianapolis: Hackett, 1983.
- . 1986. "What are Logical Notions?" *History and Philosophy of Logic* 7: 143–154.
- Wallace, J. 1964. *Philosophical Grammar*. Stanford University Ph.D. Ann Arbor, Mich.: University Microfilms Ltd, 1969.
- Westerståhl, D. 1989. "Quantifiers in Formal and Natural Languages." In *Handbook of Philosophical Logic*, vol. 4, edited by D. Gabbay and F. Guenther. Dordrecht: Reidel.
- Wiggins, D. 2001. *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.
- Williamson, T. 1987/88. "Equivocation and Existence." *Proceedings of the Aristotelian Society* 87: 109–127.
- . 1999a. "Existence and Contingency." *Proceedings of the Aristotelian Society, Supplementary Vol. 73*: 181–203. Reprinted with printer's errors corrected in *Proceedings of the Aristotelian Society* 100: 321–343.
- . 1999b. "A Note on Truth, Satisfaction and the Empty Domain." *Analysis* 59: 3–8.
- . 2003a. "Everything." *Philosophical Perspectives* 17: 415–465.
- . 2003b. "Understanding and inference." *Proceedings of the Aristotelian Society, Supplementary Vol. 77*: 249–293.

RALPH WEDGWOOD

The Refutation of Expressivism

How should we set about the task of explaining the meaning of normative statements—that is, of statements about what *ought* to be the case, or about what people *ought* to do or to think? (As I am using the term, a “statement” is just the speech act that is performed by the sincere utterance of a declarative sentence. So a “normative statement” is just the speech act, whatever exactly it may be, that is performed by the sincere utterance of a declarative sentence involving a normative term like ‘ought.’ I shall use the term ‘judgment’ to refer to the type of mental state that is expressed by a statement; so a “normative judgment” is just the type of mental state, whatever exactly it may be, that is expressed by a normative statement.)

In this paper, I shall consider a certain well-known approach to the task of giving an account of the meaning of normative statements. This is the approach that is based on an *expressivist* account of normative statements—the approach whose most distinguished exponents in recent years have been Simon Blackburn and Allan Gibbard.¹ I shall argue against the expressivist approach, and in favor of the rival *truth-conditional* or *factualist* approach.

An earlier version of this paper was presented at the conference on truth at the Central European University in April 2005. I should like to thank that audience for their helpful comments. A slightly modified version of this paper has already been published; see R. Wedgwood, *The Nature of Normativity* (Oxford: Clarendon Press, 2007), chap. 2.

¹ See especially Gibbard 1990 and 2003, Blackburn 1993 and 1998.

1 Expressivism and non-cognitivism

According to an expressivist account of normative statements, the fundamental explanation of the meaning of normative statements, and of the sentences that are used to make those statements, is given in terms of the type of *mental state* that the statements made by uttering those sentences *express*. That is, the fundamental explanation of the meaning of these statements and sentences is given by a *psychologistic semantics*.² According to a plausible version of the principle of *compositionality*, the meaning of a sentence is determined by the meaning of the terms that it is composed out of, together with the compositional structure of the sentence (perhaps together with certain features of the context in which that sentence is used). So assuming this version of the compositionality principle, this expressivist approach will also give an account of the particular terms involved in these sentences in terms of the contribution that these terms make to determining what type of mental state is expressed by sentences involving them.

I shall suppose that all such expressivist accounts aim to conform to a basic non-circularity constraint.³ Thus, according to these expressivist accounts, the fundamental explanation of the meaning of normative statements, such as statements of the form “I ought to φ ,” must not identify the mental state that is expressed by this statement simply as *the belief that one ought to φ* (or the feeling or the sentiment that one ought to φ), or anything of that sort. This mental state must be identified without using any normative terms like ‘ought’ within the scope of propositional attitude ascriptions of any kind. Otherwise, we would be presupposing what we are seeking to give an account of—namely, what it is for a thinker to have such normative attitudes.

In fact, many expressivist accounts conform to a yet stronger constraint: they seek to give their fundamental explanation of the meaning of normative statements *in wholly non-normative terms*. So they not only avoid using normative terms in any way that would effectively amount to presupposing what it is for a statement to be a normative statement, or what it is for a judgment to be a normative judgment: they insist on

² I borrow this term from Rosen (1998, 387).

³ The sort of non-circularity constraint that I have in mind is the constraint that is articulated by Christopher Peacocke (1992, chap. 1).

banishing all normative terms from the metalanguage altogether. The main reason for this is that the proponents of expressivist semantics usually aspire to give an account of normative statements that is wholly compatible with a *naturalistic* metaphysics, according to which all of our thought and discourse can ultimately be satisfactorily explained purely in the terms that are characteristic of natural science.⁴

The main rival to the expressivist approach is a broadly *truth-conditional* or *factualist* approach. According to the truth-conditional or factualist approach, the fundamental explanation of the meaning of the term in question must essentially involve the idea that if a declarative sentence involving the term has any content at all in a given context, then that content involves⁵ a *proposition*, where it is an essential feature of any proposition that it is *truth-apt*—that is, apt to be either *true* or *false*. In this way, then, this proposition gives the *truth conditions* of the sentence in the context in question: the sentence is true in that context if, and only if, the corresponding proposition is true.⁶ It is often assumed that there is a close link—perhaps identity, or at least one-to-one correspondence—between *true propositions* and *facts*; this is why I shall also describe the truth-conditional approach to the semantics of normative terms as a “factualist” approach.

In saying that the expressivist approach is a “rival” of the truth-conditional approach, I do not mean to say that expressivists must *deny* that the content of any of the sentences involving the term in question involves a proposition that is apt to be true or false. The point is just that according to the expressivist, the *fundamental* explanation of the term’s meaning need say nothing about such propositions’ being involved in the contents of the sentences in which this term appears: this fundamental account should be given in strictly psychologicistic terms,

⁴ The importance of these two constraints for expressivists is rightly stressed by Rosen (1998, 388).

⁵ I say that according to the truth-conditional or factualist approach, the content of the sentence “involves” a proposition to allow for accounts according to which the content of the sentence also involves *other* elements—such as a Fregean Thought (if that is not to be identified with the proposition itself), or the aspects of meaning that Dummett (1981, 1–7, and 1993, 38–41) called “tone” and “force.”

⁶ I do not mean to imply that the truth-conditional approach must deny the possibility of “truth-value gaps”—that is, propositions that are neither true nor false. It may be possible for a proposition to be “apt” to have a truth-value (that is, the proposition might have truth conditions), even if in some circumstances it lacks any definite truth-value.

without mentioning these sentences' having truth conditions or propositions as part of their content.

Typically, expressivism has tended to go along with a *non-cognitivist* account of the normative judgments that are expressed by normative statements: until recently, all non-cognitivists have been expressivists, and all expressivists have been non-cognitivists. Nonetheless, these two views should in principle be distinguished. *Non-cognitivism* is a view about the nature of normative judgments. Specifically, it is the view that these normative judgments are not "cognitive" mental states, like ordinary beliefs (of the sort that under favorable conditions could constitute knowledge), but "non-cognitive" mental states of some kind, like desires, preferences, emotions, intentions, or the like.

The reason why the two views tend to go together seems to be this. First, suppose that expressivism is false. Then, it is widely assumed, the correct explanation of the meaning of normative statements will be some sort of truth-conditional or factualist account. In that case, the sentence that is used to make any normative statement has a proposition as at least part of its content; and reference to this proposition can play a significant role in an explanatory account of the nature of normative discourse. But then it is hard to see what could make it impossible for this propositional content to be the object of the full range of propositional attitudes—including not only the non-cognitive attitudes (like intention, hope, desire, and the various types of emotion), but also the cognitive attitudes (such as the type of attitude that is involved in ordinary beliefs). If it is possible to have such a cognitive attitude towards such a propositional content, then it is also hard to see why it should be impossible to express that attitude by means of making the corresponding normative statement.

Admittedly, this truth-conditional or factualist account does not imply that the mental state that is *actually normally* expressed by normative statements is a cognitive state.⁷ But it does at least make it

⁷ Recently, Mark Kalderon (2005) has investigated the possibilities of a *fictionalist* approach, according to which normative sentences do indeed have essentially truth-apt propositions as their contents, but the sincere acceptance of such a sentence does not normally consist in our *believing* the content of that sentence, but rather in having some *non-cognitive* attitude instead. In this way, Kalderon aims to exploit the fact that a factualist semantics for normative statements is at least logically consistent with non-cognitivism about the nature of normative judgments.

plausible that the burden of proof should be on someone who claims that the mental state normally expressed by the sincere utterance of a declarative sentence is something other than belief in the propositional content of that sentence. Moreover, I believe that it can be shown (although I shall not be able to show it here) that some of the best-known arguments for non-cognitivism (that is, for the view that normative judgments cannot be cognitive states) are unsound. So if, as I shall argue here, expressivism is false, we have good reasons for concluding that non-cognitivism is false as well.

Conversely, suppose that cognitivism is true. Then there is no sense in which normative judgments are any less “cognitive” than ordinary beliefs. There are strong reasons for thinking that such ordinary beliefs or cognitive states have propositions as at least part of their contents, where reference to such propositional contents is capable of playing a genuinely explanatory role in accounting for various features of the beliefs in question. But if the contents of normative judgments involve such propositions, it will surely seem irresistible to explain the meaning of normative statements at least partly in terms of these propositions. Thus, if non-cognitivism is false, it seems overwhelmingly plausible that expressivism will also be false.⁸ For these reasons then, then, it is reasonable to assume that expressivism and non-cognitivism stand or fall together. For the rest of this paper, however, my arguments will focus entirely on expressivism, not non-cognitivism.

At the beginning of this paper, I said that most distinguished recent exponents of expressivism—the view that I am planning to attack here—are Simon Blackburn and Allan Gibbard. But any attribution of views to Blackburn and Gibbard has to proceed carefully, on the grounds that both of these philosophers have recently adopted a “quasi-realist” program, seeking to show that they can accept practically all of the theses that were formerly thought to be definitive of “moral realism.” In the remainder of this section, I shall briefly defend my claim that the sort of expressivism propounded by Blackburn and Gibbard is indeed a rival of the truth-conditional or factualist approach that I have characterized here.

⁸ See Horgan and Timmons 2000 for a cognitivist form of expressivism (or as they prefer to call it, “nondescriptivism”).

These philosophers' quasi-realist program has two main elements. First, there is an element that has largely been developed by Blackburn, which involves arguing that expressivists need not deny such things as that normative statements can be *true* or *false*, or that normative terms stand for normative *properties*, or that true normative statements correspond to normative *facts*, or that ordinary people *believe* many normative *propositions*. The principal way in which Blackburn seeks to argue for this is by insisting on a relatively *minimalist* interpretation of such notions as "truth," "properties," "facts," "belief" and "propositions."⁹ For example, according to this minimalist interpretation, to claim that it is "true" that genocide is wrong is just to claim that genocide is wrong—or at least it follows immediately from the very definition of "true" that these two claims are equivalent. Similarly, to say that genocide has the "property" of being wrong is also just to say genocide is wrong; to say that there is such a fact as the fact that genocide is wrong is, once again, just to say that genocide is wrong. The claim "There is such a proposition as the proposition that genocide is wrong" is guaranteed to be true by the mere fact that the embedded sentence 'genocide is wrong' is a complete meaningful sentence; and the claim, made about a thinker *S*, "*S* believes that genocide is wrong" is true if and only if *S* has an attitude of the very same kind that one would normally express by making a sincere statement of the form "Genocide is wrong."

Even though Blackburn makes all of these claims, however, he still holds that the notions of normative propositions, normative facts or truths, and normative properties play no real *explanatory role* within the fundamental account of the meaning of normative statements. The fundamental account is still purely psychologistic: that is, it is couched entirely in terms of the mental states ("normative judgments" or "normative attitudes") that those statements express. Moreover, these

⁹ See especially Blackburn 1998, 317–319, and 1993, 3–6. As is well known, a minimalist theory of truth needs to be very carefully stated to avoid the Liar Paradox (see e.g. Halbach 2001). In fact, the same is true of minimalist theories of properties: we have to avoid being committed to the existence of the property of *being a property that does not instantiate itself*, and to the general claim that something instantiates the property of being *F* if and only if it is *F*—since taken together, these two commitments entail that this property of *being a property that does not instantiate itself* instantiates itself if and only if it does not instantiate itself. On this point, see Bealer 1982, chap. 4, esp. § 26.

notions of normative propositions, facts, and properties also play no real explanatory role within the expressivist's account of the nature of these normative mental states: that account is given in other terms (for example, most contemporary expressivists choose to account for the nature of these mental states in broadly functionalist terms). So Blackburn's minimalist quasi-realism does not prevent him from being an expressivist, or from rejecting the rival truth-conditional approach, as I have characterized them.

The second element in these philosophers' quasi-realist program has mainly been developed by Allan Gibbard. This is Gibbard's "natural constitution claim." There is a separate version of this claim involving each normative concept. For example, consider the concept that Gibbard expresses by means of the phrase 'the thing to do' (which in many contexts will be equivalent to such phrases as 'the right thing to do' or 'the thing that one ought to do'). The version of this "natural constitution claim" involving this concept would be the claim that there is some natural property *N* such that, necessarily, any action is the thing to do just in case the action has property *N*. As Gibbard argues, this claim follows, at least in the most widely accepted modal logic,¹⁰ from two plausible assumptions: (i) the assumption that the normative *strongly supervenes* on the natural—that is, necessarily, if something is "the thing to do," then it is impossible for there to be anything that is exactly like that first thing in all natural respects without also being exactly like that first thing in normative respects as well (and so itself being the "thing to do" as well); (ii) a relaxed conception of properties according to which there is a huge plethora of properties, with at least one property for every way of mapping each possible world *w* onto a subset of the entities that exist at *w*.¹¹ If, as Gibbard argues, every competent user of normative terms is committed to these two assumptions, then every competent user of these terms is also committed to accepting the "natural constitution claim," and in that sense accepting that there is a naturalistic property that the normative term "signifies."

¹⁰ The claim does not in fact follow from these assumptions in any alethic normal modal logic weaker than S5. I explore the significance of this point in Wedgwood 2007, chap. 9.

¹¹ See especially Gibbard 2002, and 2003, chap. 5.

In this way, it seems to be a crucial part of Gibbard's account that the fundamental explanation of the meaning of normative statements *quantifies* over naturalistic properties, and includes the thesis that every competent speaker is committed to the "natural constitution claim" that *there is* a natural property such that something is "the thing to do" if and only if it has that property. However, it still seems that this fundamental explanation does not actually need to *refer* to any *particular* property as the property that the normative term signifies. In that sense Gibbard's semantics is not really a referential semantics for these normative terms. Every competent speaker is committed to claiming that the term 'the thing to do' "signifies" a particular natural property *N*, but as Gibbard (2003, 115–116) puts it, such claims "are not purely matters of linguistic fact, or of linguistic and psychological fact combined."

Thus, according to Gibbard, the meaning of the normative term can be explained without referring to any particular property as the property that the term signifies. In this sense, then, Gibbard's semantics remain crucially different from a truth-conditional semantics, according to which the fundamental explanation of the meaning of normative statements must refer to a particular property or relation in giving the truth conditions for these normative statements. Thus, Gibbard does not accept that there is any particular property determined purely by the *meaning* of the relevant normative term (such as 'ought' or 'the thing to do') as the property that that term stands for; his account must still be distinguished from a truth-conditional account.

2 Geach's Fregean problem

In arguing against expressivism, I will build on the objections that other philosophers have made against expressivism. One of the main objections that I will build on here is the broadly Fregean argument that P.T. Geach used against an earlier generation of expressivists.

A memorable statement of the views of these earlier expressivists was given by A.J. Ayer (1946, chap. 6). Ethical concepts, according to Ayer, "are mere pseudo-concepts." The role of ethical symbols in language is not to add anything to the "factual content" of statements in which they appear, but simply to show that the utterance of the

sentence “is attended by certain feelings in the speaker” (Ayer 1946, 107). The function of an ethical word is purely “emotive.” “It is used to express feeling about certain objects, but not to make any assertion about them” (Ayer 1946, 108). ‘Stealing is wrong’ simply expresses a feeling of moral disapproval towards stealing; it does not offer a description of stealing, or ascribe any property to it. This is why normative statements lack truth conditions, and cannot be described as true or false.

This relatively crude form of expressivism was decisively refuted by Geach (1972, 250–269), in an argument that was inspired by a famous point that Frege (1977, 45–48) made about negation. Some normative statements are effected by the utterance of sentences in which normative terms have largest scope; and it may be plausible to regard these statements as expressions of emotion or as commands or prescriptions. But sentences containing normative terms can also be *embedded* within the scope of sentential operators of all kinds.

For example, sentences containing normative terms like ‘ought’ and ‘wrong’ occur in the antecedents of conditionals, or as disjuncts in disjunctions, or as the objects of propositional attitude ascriptions. Normative terms also occur in non-declarative sentences, the utterance of which does not count as a statement but as a speech act of some other kind, such as a question. In many of these contexts, these occurrences of normative terms simply cannot be seen as expressing the speaker’s feelings in the way that Ayer describes. If I say “Rape is wrong,” I may well be expressing moral disapproval of rape. But if I say “Helen believes that abortion is wrong,” or “If gambling is wrong, then encouraging people to gamble is also wrong,” or if I ask “Is the death penalty really always wrong?,” I am not expressing moral disapproval of abortion or gambling or the death penalty. Yet the word ‘wrong’ is plainly used in the same sense throughout. It would be unbearably *ad hoc* to claim that normative terms are systematically ambiguous, depending on whether or not they have largest scope in the sentences in which they occur. So Ayer’s account of the meaning of normative terms is unacceptable as it stands. An adequate account of the meaning of normative terms must explain how they can figure, without a shift of meaning, both in statements in which they have largest scope, and embedded in subsentences of complex utterances in which they do not have largest scope.

Moreover, it seems to me, we should not only require that normative terms can occur, without a shift of meaning, both in embedded contexts and elsewhere; we should also require a uniform interpretation of the sentential operators within whose scope they are embedded. It would seem to me even more intolerably *ad hoc* to claim that the logical operators are systematically ambiguous, depending on whether or not the sentences embedded within their scope contain normative language. ‘If’ may be tricky, but surely it is not quite as tricky as that! To adopt a point of Elizabeth Anscombe’s (1963, 58), if we are willing to tolerate a special “normative conditional,” then why not tolerate a special “mince pie conditional” (whose meaning has to be explained by appealing to special facts about sentences concerning mince pies)? We should respect the intuition that, in the statement “If gambling is wrong, encouraging others to gamble is also wrong,” ‘if’ is used in a perfectly ordinary sense.

Since they can be embedded within the logical operators, normative terms can also occur in logical inferences of all the customary kinds. An account of the meaning of a logical operator must supply an account of the validity of the customary forms of inference in which it appears; so if we insist on a uniform account of the meaning of ‘if,’ we must also insist on a uniform account of the validity of all these forms of inference. In the case of ‘if,’ the central example is *modus ponens*. Take any inference of the form: “If *A* then *B*; but *A*: Hence *B*.” This inference is valid in exactly the same way regardless of whether or not ‘*A* or *B*’ contains normative language. Expressivists must not only explain how normative terms can occur, without shifts of meaning, both in embedded contexts and elsewhere; they must also explain how the meaning of normative terms and of the logical operators allows all the customary forms of inference to be valid, in exactly the same way as they usually are, regardless of whether or not they contain normative language.¹²

¹² Oddly enough, many philosophers who have written about Geach’s Fregean problem seem open to the postulation of a special “normative conditional,” and of special forms of inference involving this special conditional. Dummett’s attempt (1981, 327–354) to solve the problem is of this kind, as is Simon Blackburn’s first attempt (1984, 189–195) at the problem. Some acute critics of Blackburn, such as Bob Hale (1993), are willing to consider such attempts in great detail. My complaint about these attempts to circumvent Geach’s problem is essentially the same as Susan Hurley’s objections (1989, 180–185) to Dummett and Blackburn.

This is a strong requirement to impose. But we should not assume straight off that it can only be met by adopting a truth-conditional semantics for normative statements. Indeed, it seems to me that the problem, at least in the form in which it has been stated so far, is adequately solved by Gibbard's recent version of expressivist semantics.

3 Gibbard's solution to Geach's Fregean problem

Ayer maintained that the meaning of a normative statement consisted in the type of emotion that it expressed. In an essentially similar way, Gibbard maintains that the meaning of a normative statement consists in its expressing a special type of mental state that he calls a "normative judgment." Ayer did not offer any analysis of emotions. By contrast, Gibbard offers an elaborate account of the nature of normative judgments—an account that is explicitly designed to provide a solution to Geach's Fregean problem.

In his earlier work, Gibbard offered a broadly *functionalist* account of normative judgments.¹³ In his most recent work (Gibbard 2003, chap. 3), he does not reject this sort of functionalism, but he no longer explicitly relies on it. Instead, his idea is to give an account of the nature of normative judgments by using two basic psychological notions. The first is the idea of having a *plan*, where this plan may be a *contingency plan*—a plan about what to do in some circumstance when one regards it as possible but not certain that one will at some time be in that circumstance—or even a purely *hypothetical plan*—that is, a plan about what to do in a circumstance that one knows full well one will never be in. Thus, according to Gibbard (2003, 53), I could have a hypothetical plan about what to do in the circumstance of being Julius Caesar on the brink of the Rubicon in 49 BC. The second psychological notion that he relies on is the notion of the mental state of *disagreeing with* (or as he also puts it, *rejecting* or *ruling out*) an action or a mental state.

According to Gibbard, the content of a normative judgment is determined by the actions and other mental states that it disagrees with—and especially by the *plans* and *attitudes of disagreement* that it disagrees with. The type of normative judgment that Gibbard analyses

¹³ For this earlier account, see Gibbard 1990, 101–102, and 1992.

in most detail consists of judgments about what courses of action are, and which courses of action are not, “the thing to do.” According to his analysis, what is essential to the judgment that in circumstances C_1 , φ -ing is the thing to do is that this judgment consists in an attitude of *disagreeing* with any course of action that involves not φ -ing in circumstances C_1 . Now, the judgment that φ -ing is “the thing to do” in certain circumstances could also be expressed by the statement that φ -ing is *required* in those circumstances. In this way, this judgment contrasts with the judgment that φ -ing is *permissible* (but perhaps not required) in certain circumstances. According to Gibbard, what is essential to the judgment that it is permissible to φ in certain circumstances C_2 is that this judgment consists in an attitude of *disagreeing with the attitude of disagreeing with* (as Gibbard puts it, an attitude of “permitting”) the action of φ -ing in circumstances C_2 .

In order to state this view more precisely, Gibbard introduces the idea of a *hyperdecided* overall mental state (a “hyperstate,” as I shall call it for short).¹⁴ Such a hyperstate would include the following two elements. First, it would include a complete consistent set of *beliefs* about ordinary factual matters. Second, it would include a *hyperplan*—roughly, a complete and consistent plan about what to do and what to think for every conceivable circumstance.

To make the notion of a hyperplan more precise, we will first need to identify a certain category of mental states. According to Gibbard (1990, 57 and 72–76), this category of mental states consists of those states that are susceptible to “normative governance”—that is, the states that one can have as a direct result of planning to have those states. According to Gibbard, these mental states include beliefs and plans; they also include the attitude of *disagreeing with* a course of action, and, for every mental state that they include, they include the attitude of disagreeing with (or rejecting or ruling out) that mental state.

We can now explain the sense in which these hyperplans must be *complete*: every one of these hyperplans must contain, for every possible circumstance, a plan to do one of the alternative actions that are available in that circumstance, and for every mental state in the relevant category that is available in that circumstance, either a plan to have or a plan not to have that mental state. Moreover, for every possible

¹⁴ I am grateful to Gibbard for helping me to understand his views.

circumstance, and every course of action and every mental state in the relevant category, the hyperstate must include a plan either to have the attitude of *disagreeing with* that course of action or that mental state in that circumstance, or else to have the attitude of *permitting* (that is, *disagreeing with disagreeing with*) that course of action or mental state in that circumstance.

We can also explain the sense in which this hyperstate must be *consistent*: first, all the beliefs that it contains must be logically consistent with each other; secondly, for every possible circumstance, it must be logically possible to realize everything that the hyperplan prescribes for that circumstance; and thirdly, if the hyperstate contains a given mental state, or a plan to have a given mental state in a given circumstance, it will never also contain a plan to *disagree with* having that mental state in that circumstance.

Now we can continue to identify an atomic *normative* judgment either with an attitude of disagreeing with an action or mental state, or with an attitude of permitting (disagreeing with disagreeing with) an action or mental state, in the way that I explained above. We can say that an atomic normative judgment disagrees with a hyperstate if and only if the hyperstate contains a plan to disagree with this attitude. If a judgment does not disagree with a hyperstate, we can say that the judgment “allows” that hyperstate. For example, consider the judgment that the thing to do now is to turn to the left. We can identify this judgment with the attitude of disagreeing with the action of *not* turning left. So this judgment allows all and only those hyperstates that do *not* include a plan to permit (that is, disagree with disagreeing with) the action of not turning left.

We can now explain the content of logically complex judgments in the following way. Let us interpret the disjuncts of a disjunctive judgment as themselves judgments—namely, the judgments that one would be expressing by uttering the disjuncts of the disjunctive sentence that can be used to express the disjunctive judgment in question; and similarly for the other logical operators. Then we can say that a disjunctive judgment allows a hyperstate if and only if at least one of its disjuncts allows that hyperstate; a negation allows a hyperstate if and only if the judgment of which it is the negation does not allow that hyperstate; and so on.

Moreover, we can also give a quite straightforward account of the notions of logical consequence and logical consistency: a conclusion

logically follows from a set of premises if and only if every hyperstate that is allowed by all the premises of the inference is also allowed by the conclusion of the inference; and a set of judgments is logically inconsistent if and only if there is no hyperstate that is allowed by all members of that set.

The crucial feature of this account is that it offers an absolutely *uniform* account of the logical operators, and of the validity of the customary forms of inference—an account that applies both to cases in which the premises of the inference are normative judgments and to cases in which they are ordinary factual judgments. In the case of purely factual non-normative judgments, the only relevant part of the hyperstate is the set of beliefs; since this set of beliefs is complete, it in effect mimics a possible world. So in the case of purely factual non-normative judgments, Gibbard's semantics coincides with a familiar sort of possible-worlds semantics. It is only when normative statements are involved that we need to take account of the fact that these hyperstates involve plans as well as beliefs.

In this way, then, Gibbard appears to have produced interlocking accounts of the meaning of normative statements and of the logical operators, and of the validity of the customary rules of inference, that meet all the constraints that we have at least so far seen to be imposed by Geach's Fregean puzzle.

4 A further constraint on accounts of normative statements

There is a further element of the meaning of normative statements that we have not yet discussed. As Crispin Wright (1992, 74) would put it, normative discourse is a thoroughly *disciplined* discourse: in making normative statements, speakers aim to comply with, and are assessed or evaluated according to, certain standards of justification or warrantedness.

As Dummett (1993, 57 and 72–76) has emphasized, an utterance may be assessed or criticized in many ways—for example, as impolite, as in poor taste, or as a breach of confidence. What will be especially relevant for our purposes are the special sorts of criticism that apply to statements precisely *because* they are statements, and not because these statements belong to some wider category of communicative action,

or because they are statements of some special kind. Thus, for example, even though statements may be criticized when they are impolite, this is not especially because they are statements. All communicative actions (including questions, requests, and gestures, as well as statements) may be criticized if they are impolite. So even though statements may be criticized when they are impolite, this is not because they are statements, but because they belong to the wider category of communicative actions. On the other hand, some statements are made in special institutional contexts in which they are open to rather special criticisms. For example, when testifying in court, a statement that is made on the basis of hearsay may be criticized, whereas outside this special institutional context, there may be no objection at all to making statements on the basis of reliable hearsay evidence. Thus, the reason why courtroom testimony based on hearsay may be criticized is not simply that such testimony consists of statements, but rather that it consists of statements of this special kind.

What we are concerned with here is the special kind of assessment or evaluation that applies to statements precisely *because* they are statements. One striking feature of this special kind of assessment is that whenever a statement is open to an assessment of this kind, a precisely analogous assessment also applies to the inner mental event of judgment, which the statement expresses—that is, the judgment that the speaker could self-ascribe by using the very same sentence that he used to make the statement in question, embedded inside a phrase like ‘I judge that ...’. If according to this special kind of assessment it is unwarranted or unjustified for one to make a certain statement, then in some closely analogous way it is also unwarranted or unjustified for one to make the judgment that that statement expresses. Similarly, if a statement is justified or warranted, then in a closely analogous way the judgment that that statement expresses is also justified or warranted.

Another distinctive feature of this special kind of assessment that statements are subject to is the connection between this sort of assessment and *logic*. Of course, logic is relevant to other sorts of speech acts as well. For example, one can certainly criticize someone if he issues a set of commands that are jointly inconsistent, just as one can criticize someone if he makes an inconsistent set of statements. But the relationship between logic and statements is nonetheless importantly

different from the relationship between logic and commands. Part of the difference is that while it is a fundamental criticism of a set of commands that they are mutually inconsistent, there are plenty of situations in which some instance of the following schema is true: one would be justified in commanding that p , and one would also be justified in commanding that $\text{not-}p$. In such situations, one cannot sensibly make both commands, but one would be equally justified in making either. It seems that this situation can never arise with statements. If one is justified or warranted in stating that p , then one is not justified or warranted in stating that $\text{not-}p$, and *vice versa*.¹⁵

Moreover, there is a further connection between the status of statements as justified or warranted and *logically valid arguments*. If a given statement is justified or warranted, and one competently infers a further statement from the first statement by means of a logically valid argument, and the first statement remains justified or warranted even after one has inferred that further statement, then the further statement is also justified or warranted as well. A further connection between logically valid arguments and the standards of assessment, evaluation and criticism that apply to statements emerges with respect to *suppositional reasoning*. In addition to straightforwardly *making* a statement, one can use the sentence that would be involved in making the statement to express a *supposition*—where to suppose that p is not straightforwardly to state that p , but simply to hypothesize that p for the sake of argument. Now one can infer a conclusion, not just from a sentence that one is using to make a statement, but also from a sentence that one is using merely to express a supposition—where to “infer” this conclusion means, roughly, that one accepts this conclusion *conditionally*, given the supposition that is expressed by the first sentence. Whenever one competently infers a conclusion from a supposition by means of a logically valid argument, then one is warranted or justified in inferring that conclusion from that supposition—that is,

¹⁵ James Lenman (2003) argues that, given that inconsistency is clearly a serious failing in any set of commands, it should be possible to develop a logic for commands, which could then be used as the logic of normative statements. I believe that this approach is vitiated by the profound differences, which I have highlighted here, between the sorts of “discipline” that commands and statements are subject to.

one is justified in accepting the conclusion conditionally, given the supposition in question.¹⁶

Finally, we should note that the conditions under which statements and inferences count as warranted or justified seems to supervene on the *meaning* of the statements or sentences involved. If there are two statements or inferences that do not differ at all with respect to the meaning of the statements or sentences involved, then it seems that they also cannot differ with respect to the conditions under which they count as warranted or justified in the relevant sense.

All of these points should be accepted by expressivists as applying to normative statements, just as much as to paradigmatically factual assertions. The standards or conditions of justification or warrantedness that apply to normative statements supervene on the meaning of the statements involved; they are intimately connected, in the ways that we have examined, with the notions of logical consistency and validity, which the expressivists also regard as notions that the fundamental account of the meaning of normative statements must explain; and the standards or conditions of justification or warrantedness that apply to a normative statement are precisely analogous to those that apply to the corresponding normative judgment, which is what the expressivist appeals to in order to explain the meaning of the normative statement. Taken together, all these points make the following conclusion highly plausible: the fundamental explanation of the meaning of a normative statement must provide some account of these conditions or standards of justification and warrantedness.

Many expressivists seem quite happy to embrace this conclusion. In particular, Gibbard spent a large part of his earlier book (Gibbard 1990) characterizing the standards of justification and warrantedness

¹⁶ This formulation of this connection between logically valid arguments and the standards of warrantedness and justification that apply to statements involves the notion of inferring *competently*. I cannot give a full discussion here of what such competence involves. But here is a quick suggestion. Suppose that one has a *disposition* in virtue of which one tends to respond to one's considering a logically valid inference by making that inference. (Of course, one's disposition will be imperfect: one will be liable to fail to respond to some of the logically valid inferences that one considers; but this does not prevent one from having a disposition of this sort.) Then, if one manifests this disposition by making a valid inference, this manifestation of the disposition will count as a case of "inferring competently" in the relevant sense.

to which normative judgments are subject. According to Gibbard, these standards consist chiefly in the consistency of one's normative judgments both among themselves, and with one's higher-order norms (or "epistemic stories") about the best way to go about forming such normative judgments (1990, 193). He also argues that one must accord a certain "fundamental authority" to the normative statements of other speakers (180), especially of other speakers in one's own community (203): we must try to avoid paying the "price" of regarding others as bad judges and excluding them from discussion (197). In effect, in Gibbard's view, warranted normative judgments arise from a process of striving for ever greater consistency, both among one's own judgments, and within the community at large.

However, it is not clear that it is enough for an account of the meaning of normative statements simply to *enumerate* the standards of justification and warrantedness that these statements are subject to. Consider an agent who is agonizing about a normative question. For example, suppose that she is agonizing over the question of whether she ought to inform the police about a friend's criminal activities. In agonizing about the question, she is striving to reach an answer to this question that is justified and warranted. But why should she bother agonizing about this? What is the point of going to so much trouble? What would be so bad about reaching an answer to the question that is not justified or warranted?

Suppose that Gibbard's account of these standards of justification and warrantedness is basically correct. Why should the agent strive so hard to achieve the sort of intrapersonal and interpersonal consistency that is such a crucial element of meeting these standards of justification and warrantedness? It surely cannot just be that we simply have a bizarre *fetish* for logical consistency, thinking that it makes for a *prettier overall pattern* in our statements or mental states than inconsistency does.

Several philosophers, including both Simon Blackburn (1992, 951) and Walter Sinnott-Armstrong (1993, 301–302), have inquired how exactly Gibbard can explain what is *bad* about making an inconsistent set of judgments according to his theory.¹⁷ Gibbard's most recent response to this inquiry is as follows:

¹⁷ Rosen (1998, 391–392) asks the same question of Blackburn's account of the logic of normative statements.

A set of judgments is *consistent* if there is a hyperstate that every judgment in that set allows. It is *inconsistent* otherwise: it is inconsistent if every possible hyperstate is ruled out by one or another of the judgments in the set. If, then, my judgments are inconsistent, there is no way I could become fully opinionated factually and fully decided on a plan for living—no way that I haven't, with my judgments, already ruled out. (Gibbard 2003, 59)

But suppose that my judgments, taken together, rule out every possible hyperstate (so that there is no way that I have not already “ruled out” in which “I could become fully opinionated factually and fully decided on a plan for living”). What is so bad about that? After all, as Gibbard (2003, 54) concedes none of us will ever be in such a “hyperstate,” and indeed “trying to approach this ideal would be a waste.” At some points, Gibbard seems to suggest that the importance of consistency in one's plans is connected with the need for one's plans to offer one guidance about what to do. But suppose that I have inconsistent plans about what to do in the position of being Julius Caesar on the brink of the Rubicon in 49 BC. These plans will then be incapable of providing me with any guidance on what to do in that situation. But so what? I know full well that I will never be in that situation. What is so bad about having inconsistent fantasy plans about what to do in that situation?

In the next section, I shall argue that any acceptable solution to this problem will in effect be incompatible with expressivism. In this way, it turns out, expressivism is incapable of providing a satisfactory account of normative discourse.

5 Expressivism defeated

Even without giving an account of the point or purpose of conforming to the standards of justification or warrantedness that apply to normative statements, we can identify certain features that this point or purpose must have. Let us return to the example of the agent who is agonizing about whether or not she ought to inform the police about her friend's criminal actions. It seems reasonable to assume that there must be some *desirable property*, which a judgment could have, such that in agonizing about this question, our agent is striving to reach a

judgment about this question that has this desirable property. (I shall consider an objection to this assumption at the end of this section.) Similarly, in agonizing about whether or not to accept the statement that she ought to inform the police about her friend, she is striving to ensure that she accepts that statement only if that statement has a certain analogous desirable property.

To give this desirable property a label, let us say that a statement that has this desirable property is a “winning” statement, and a judgment that has the analogous desirable property is a “winning” judgment. Let us also say that if the statement that would be performed by uttering a given sentence (in a given context) is a winning statement, then the sentence in question is a “winning” sentence (in that context). The point or purpose of conforming to these standards of warrantedness or justification is to ensure that one makes only “winning” statements and “winning” judgments. This does not mean that conforming to these standards *guarantees* that one makes only winning statements and judgments. It may be possible, if one is unlucky, for a judgment to be justified even if it is not in fact a winning judgment. The point is rather (to put it roughly) that conforming to these standards of justification is the *means* that one uses in order to achieve the goal of making only winning statements and judgments.

As I have already argued, the fundamental explanation of the meaning of a normative statement must give some account of the specific standards of justification or warrantedness that the statement is subject to. But as we have seen, the very point and purpose of these standards is to ensure that one makes only winning statements. So it also seems plausible that this fundamental explanation of the meaning of the statement must also give some account what it would be for this statement to be a winning statement.

Whatever exactly this desirable property of being a “winning” statement may be, the goal of making only winning statements must be served by conforming to these standards of justification and warrantedness—and in particular it must be served by conforming to standards that have the precise connection to logic that I have described.

One of the connections between logic and these standards of justification and warrantedness is the principle that if one is justified or warranted in making a given statement, then one is not justified or warranted in making any statement that is logically inconsistent with the first statement. So it seems that this desirable property of being

a “winning” statement must have an analogous connection to logical inconsistency. That is, no statement that is logically inconsistent with a winning statement can itself be a winning statement.

Moreover, the attitude of *inferring* the conclusion of an argument (that is, accepting the conclusion of the argument *conditionally*, given the supposition of the argument’s premises) can presumably also be a winning attitude. The obvious suggestion to make here is the following. If one infers a conclusion from certain premises, and this attitude of inferring that conclusion from those premises is a winning attitude, then *if* each of the premises is a winning sentence, the conclusion of the argument is also a winning sentence. That is, if the attitude of inferring the conclusion from the premises is a winning attitude, then the property of being a winning sentence is *preserved* from the premises to the conclusion. Since a relevantly competent reasoner is always warranted in accepting a logically valid inference, it seems that the property of being a winning sentence is always preserved from the premises to the conclusion of any logically valid argument.

If this is true, then we can derive numerous further features of this desirable property of being a winning sentence: a disjunctive sentence has this property if and only if at least one of its disjuncts has this property; a sentence in which negation is the dominant operator has this property just in case the sentence embedded inside the negation operator does not have this property; and so on. Moreover, there is in fact an extremely easy way of supplementing Gibbard’s semantics for normative statements so that it can explain all these features of the property of being a winning statement. Let us just say that some of the “hyperstates” that Gibbard invokes are *designated* states, while the other such hyperstates are not designated. Then we could say that an atomic judgment is a winning judgment just in case it allows all of the designated hyperstates.¹⁸ We could give the semantics of the logical operators in the following simple way: the negation of a judgment counts as a winning judgment if and only if the judgment of which it is the negation does not count as a winning judgment; a disjunctive judgment counts

¹⁸ I am allowing that *several* hyperstates may count as “designated” to accommodate “Buridan’s ass” situations, where all the designated hyperstates contain an attitude of *permitting both* going left *and* going right, and so some designated hyperstates will include a plan to go left, while the other designated hyperstates include a plan to go right.

as a winning judgment if and only if at least one of its disjuncts is a winning judgment; and so on.

There is one final feature of this property of being a winning sentence that I should like to highlight. Suppose that one makes a statement; then it seems that one is thereby committed to accepting that the statement that one has made is a winning statement—that is, not merely that the statement is justified or warranted, but that the statement has achieved the point or purpose of conforming to the relevant standards of justification or warrantedness—and so the sentence that one utters in making that statement is itself a winning sentence. Conversely, suppose that one accepts that a particular sentence is a winning sentence; then it seems clear that one is thereby committed to accepting the statement that one could make by uttering that sentence. In short, the sentences *s* and ‘*s* is a winning sentence’ are equivalent, in the sense that the statement made by uttering one of these sentences commits one to the statement that one could make by uttering the other.

The identity of this property of being a “winning” sentence should by now be clear. We have established that this property has the following features: an account of the meaning of a sentence must include some account of the condition that has to be met if the sentence is to have this property; if a sentence has this property, then the statement that is made by uttering this sentence will achieve the point or purpose of conforming to the distinctive standards of justification or warrantedness that apply to statements; if a sentence *s* has this property, then no sentence logically inconsistent with *s* has this property; this property is preserved in all logically valid inferences (so that a disjunctive sentence has this property if and only if at least one of its disjuncts has this property; a negative sentence has this property if and only if the negated sentence does not have this property; and so on); and accepting that a sentence *s* has this property commits one to accepting the statement that one could make by uttering that very sentence.

It is surely highly plausible that the only property of sentences that has all these features is *truth*.¹⁹ To be a “winning” sentence is just to be

¹⁹ Suppose that the equivalence of the sentences *s* and ‘*s* is a winning sentence’ warrants all of the Tarski-style biconditionals (such as “‘Snow is white’ is a winning sentence if and only if snow is white,” and the like). Then if there were any two properties that have all these features, these properties would be exemplified by exactly the same sentences—which surely makes it seem more plausible that they are not in fact two distinct properties after all.

a *true* sentence; and to be a “winning” statement is just to be a *correct* statement, in the sense in which it is plausible to say that a statement is correct if and only if the sentence that one uttered in making that statement is true.

To say that truth is the only property of sentences that has all these features is not to embrace any sort of minimalist or deflationist conception of truth.²⁰ First, since these features include the feature of being the point or purpose of conforming to the distinctive standards of justification or warrantedness that apply to statements, they already go beyond the “platitudes” usually appealed to by such minimalist or deflationist accounts of truth. Indeed, my suggestion that it is one of the distinguishing marks of truth that truth is the “point” or “purpose” of conforming to the distinctive standards of justification or warrantedness that apply to statements has much more in common with Dummett’s idea (1993, 42–52) that the “root of our concept of truth” is our grasp of what it is for a belief or an assertion to be correct, or with Wiggins’ idea (1989, 147) that “truth is the primary dimension of assessment for beliefs,” than with any ideas in the minimalist or deflationist tradition. Secondly, to say that truth is the only property that has these features is not to say that these features exhaust the nature or essence of truth; it may well be that truth has some deeper nature—such as the nature that is articulated by some version of the correspondence theory of truth²¹—even if it is also the only property that has the features that I have identified.

I have already argued that, even at the most fundamental explanatory level, an adequate explanation of the meaning of a normative sentence must give some account of the conditions that must be met for the sentence to have this desirable property. Since this desirable property is in fact truth, any adequate account of the meaning of a normative sentence must somehow explain the sentence’s truth conditions. That is, the semantics of normative terms is truth-conditional, contrary to what the expressivist claims.

²⁰ For minimalist or deflationary conceptions, see Horwich 1998, and especially Field 1986 and 1994. My first attempt at developing this sort of argument against expressivism (Wedgwood 1997) appears to have been interpreted as committed to a form of minimalism by some readers (see Lenman 2003, note 47). I wish to emphasize here that this interpretation is mistaken.

²¹ I am in fact sympathetic to the version of the correspondence theory of truth that is advocated by Bealer (1982, 199–204). But I obviously cannot defend this theory here.

One serious objection to the argument that I have given in this section ought to be considered at this point. At the beginning of the argument, I assumed that if there was some further point or purpose in conforming to the standards of justification and warrantedness that apply to normative statements and judgments, then there must be some “desirable property” such that the point or purpose of conforming to these standards of justification and warrantedness is to ensure that each of one’s normative statements has this property. But perhaps there is no such desirable property that *individual* statements can have, but only a desirable property that whole *sets* of normative statements can have? Perhaps the point or purpose of conforming to these standards of justification and warrantedness is to ensure that the whole *set* of statements that one accepts has this desirable property? What justifies my assumption that there is a desirable property of *individual* statements such that the point or purpose of conforming to these standards of justification and warrantedness is to ensure that each of one’s statements has this property?

Suppose that there is a property P_1 such that the ultimate point or purpose of conforming to these standards is just to ensure that the whole set of normative statements that one accepts has property P_1 . Then of course we can just *define* a property P_2 that individual statements can have—namely, the property of being a statement that belongs to a set that has this desirable property P_1 . Now, one might still object that even if there is a set of statements each of which has this property P_2 , it does not follow that the whole set will have the desirable property P_1 . However, once we take into account some of the features that this desirable property P_1 must have, we will see that this last objection is just false: any set of statements each member of which has P_2 will itself be a set of statements that has P_1 .

First, as I have argued, these standards of justification and warrantedness imply that if one is justified or warranted in accepting a statement, then one is *not* justified or warranted in accepting the negation of that statement, and *vice versa*. So it seems that if a set that has this desirable property P_1 contains a statement s , then *no* set that contains the negation of s can have this desirable property P_1 . Secondly, it seems that there will be a *maximal* set that has this desirable property P_1 —specifically, a set that, for every statement for which one might possibly have any justification at all, contains either that statement or

its negation. Taking these two points together, it follows that there will be a *unique maximal* set of statements that has this desirable property P_1 , such that every set of statements that has P_1 is itself a subset of this unique maximal set. Finally, one is justified in accepting a set of statements if and only if one is also justified in accepting each subset of that set of statements; so it seems that a set of statements can have this desirable property P_1 if and only if each of its subsets also has this property P_1 . Thus, if a statement belongs to *any* set that has this desirable property P_1 (that is, if the statement has desirable property P_2), it will belong to this unique maximal set that has property P_1 ; and any set of such statements will be a subset of this unique maximal set, and so will itself also have property P_1 . So in fact, any set of statements each member of which belongs to some set that has this desirable property P_1 (that is, any set each member of which has property P_2) will itself be a set that has the desirable property P_1 . Hence, the aim of accepting only sets of statements that have this desirable property P_1 does not differ in any way from the aim of accepting only statements that have property P_2 : nothing could possibly promote one of these aims without also promoting the other. For this reason then, it was reasonable for me to assume that there is some desirable property, which *individual* statements can have, such that the purpose or point of conforming to the standards of justification and warrantedness that apply to normative statements is to ensure that each and every one of the statements that one accepts has this desirable property.

It seems then that my argument is sound: the meaning of normative statements is truth-conditional. The conclusion of this argument does not imply metaphysical realism with respect to normative truths: in particular, this conclusion is quite compatible with the claim that normative truths are wholly reducible to naturalistic truths. Nor does this conclusion establish the sort of semantic realism that Dummett (1993) has famously tried to raise problems for. For all that I have argued so far, the notion of truth applicable to normative statements might just be the notion that is favored by intuitionistic logicians, according to which for a statement to be true is for there to be an effective procedure that will yield a proof of it.²² All that I have argued

²² For this conception of truth, see Dummett 1978, 313–315, and Tennant 1987, 128–133.

here is that expressivism fails, and that even at the most fundamental explanatory level, any adequate account of the meaning of normative statements must ascribe truth conditions to these statements.

6 Gibbard's semantics transformed

I remarked above that there is an easy way of revising Gibbard's semantics so that it can accommodate the idea of "winning" sentences or statements—by adding the idea some of the hyperstates are, as I put it, "*designated*," while others are not. Then Gibbard could say that an atomic normative sentence or statement counts as "winning" if and only if it allows all these designated hyperstates.

Now that we have seen that the property of being a winning sentence is simply *truth*, and the property of being a winning statement is the property of being *correct*, we can reformulate this revision of Gibbard's semantics. These "designated" hyperstates are simply the *correct* hyperstates. That is, all the beliefs and plans contained in these hyperstates are correct: each of these beliefs is a belief in a truth; and each of the "hyperplans" contained in these hyperstates involves planning on doing everything that one really ought to be doing, and nothing that one ought not to be doing. When supplemented in this way, Gibbard's theory will imply that an atomic normative sentence, of the form 'One ought to φ in C ,' is true if and only if the attitude of disagreeing with the plan of not φ -ing in C is allowed by all the correct hyperstates—that is, no correct hyperstate permits the plan of not φ -ing in C . (Then we can explain the truth conditions of logically complex sentences in the standard truth-conditional way: a disjunctive sentence is true if and only if at least one of its disjuncts is true; a negative sentence is true if and only if the negated sentence is not true; and so on.) Indeed, this is, more or less, the sort of account of the truth conditions of normative sentences that I would regard as broadly speaking correct. Obviously, however, I cannot defend this sort of account here. All that I have argued for here is that it appears that expressivist theories face insuperable problems, and hence that the correct sort of semantics for normative statements must be broadly truth-conditional or factualist in form.

References

- Anscombe, G.E.M. 1963. *Intention*. 2nd edition. Oxford: Blackwell.
- Ayer, A.J. 1946. *Language, Truth and Logic*. 2nd edition. London: Victor Gollancz.
- Bealer, G. 1982. *Quality and Concept*. Oxford: Clarendon Press.
- Blackburn, S. 1984. *Spreading the Word. Groundings in the Philosophy of Language*. Oxford: Clarendon Press.
- . 1992. "Gibbard on Normative Logic." *Philosophy and Phenomenological Research* 52: 947–952.
- . 1993. *Essays on Quasi-Realism*. Oxford: Clarendon Press.
- . 1998. *Ruling Passions: A Theory of Practical Reason*. Oxford: Clarendon Press.
- Dummett, M. 1978. *Truth and Other Enigmas*. London: Duckworth.
- . 1981. *Frege: Philosophy of Language*. 2nd edition. London: Duckworth.
- . 1993. *The Seas of Language*. Oxford: Clarendon Press.
- Field, H. 1986. "The Deflationary Conception of Truth." In *Fact, Science and Value*, edited by G. McDonald and C. Wright, 55–117. Oxford: Blackwell.
- . 1994. "Disquotational Truth and Factually Defective Discourse." *Philosophical Review* 103: 405–452. Reprinted in Field 2001.
- . 2001. *Truth and the Absence of Fact*. Oxford: Clarendon Press.
- Frege, G. 1977. *Logical Investigations*. Translated by P.T. Geach and R.H. Stoothoff. Oxford: Basil Blackwell.
- Geach, P.T. 1972. *Logic Matters*. Oxford: Basil Blackwell.
- Gibbard, A. 1990. *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.
- . 2002. "Normative and Recognitional Concepts." *Philosophy and Phenomenological Research* 64 (1): 151–167.
- . 2003. *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.
- Halbach, V. 2001. "How Innocent is Deflationism?" *Synthese* 126: 167–194.
- Haldane, J. and C. Wright. 1993. *Reality, Representation and Projection*. New York: Oxford University Press.
- Hale, B. 1993. "Can There Be a Logic of Attitudes?" In Haldane and Wright 1993, 337–363.
- Horgan, T. and M. Timmons. 2000. "Nondescriptivist Cognitivism: Framework for a New Metaethic." *Philosophical Papers* 29: 121–153.
- Horwich, P. 1998. *Truth*. 2nd edition. Oxford: Clarendon Press.
- Hurley, S.L. 1989. *Natural Reasons*. New York: Oxford University Press.
- Kalderon, M. 2005. *Moral Fictionalism*. Oxford: Clarendon Press.
- Lenman, J. 2003. "Disciplined Syntacticism and Moral Expressivism." *Philosophy and Phenomenological Research* 66: 32–57.
- Peacocke, C. 1992. *A Study of Concepts*. Cambridge, Mass.: MIT Press.

- Rosen, G. 1998. "Blackburn's *Essays in Quasi-Realism*." *Noûs* 32: 386–405.
- Sinnott-Armstrong, W. 1993. "Some Problems For Gibbard's Norm-Expressivism." *Philosophical Studies* 69: 297–313.
- Tennant, N. 1987. *Anti-Realism and Logic*. Oxford: Clarendon Press.
- Wedgwood, R. 1997. "Non-Cognitivism, Truth, and Logic." *Philosophical Studies* 86: 73–91.
- . 2007. *The Nature of Normativity*. Oxford: Clarendon Press.
- Wiggins, D. 1989. *Needs, Values, Truth*. 2nd edition. Oxford: Basil Blackwell.
- Wright, C. 1992. *Truth and Objectivity*. Cambridge, Mass.: Harvard University Press.

HOWARD ROBINSON

Benacerraf's Problem, Abstract Objects and Intellect

Introduction

The target paper for the conference which was the origin of this collection was Benacerraf's "Mathematical Truth" (1973). Benacerraf's article concerns the difficulty of combining a causal theory of knowledge with the fact that, in the case of mathematics at least, the objects of our knowledge are abstract entities without causal powers. In the first section of this essay, I argue that, if this is a problem, it is not restricted to mathematics, but concerns all thought: all thought involves apprehending abstract objects, in the form of universals. The mistake is to believe that a naturalistic account of thought was ever on the cards. The ability to grasp abstract objects, whether universals or abstract particulars, like numbers, is a primitive and defining property of intellect, as Plato and Aristotle believed. In the second and third sections, I attempt to sketch out what I term a neo-Platonic account of abstract objects and their relation to thought.

1 Benacerraf's problem

1.1 The nature of the problem

Benacerraf sets out the problem that is supposed to face the realist philosopher of mathematics who wishes to be a naturalist. The argument roughly runs as follows.

- (1) On a realist—that is, a Platonist—account, numbers are abstract objects.

- (2) Knowledge of particulars [of a certain general class] involves causal interaction with them [with some instances of that class].
- (3) Naturalism entails that all events are closed under physics, so there is no causal interaction with anything non-physical.

therefore

- (4) Naturalism entails that there is no knowledge of any non-physical particulars.

therefore

- (5) Naturalism entails that, on a realist—Platonist—account of them, there is no knowledge of numbers.

Although this is not in Benacerraf, the argument might continue:

- (6) There is knowledge of numbers.

therefore

- (7) Either naturalism is false or a non-realist (e.g. a reductionist or fictionalist) account of numbers is true.

Quite how to state (2) might be controversial, but, for my expository purposes, only the general drift matters. There are two assumptions in the argument. One is that abstract objects are non-physical, the other is that they are particulars. Particulars are generally contrasted with universals, and universals would normally be regarded as abstract objects, so not all abstract objects are particulars. Numbers, however, do not seem to be the same as universals. *Prima facie*, *one* is not *unity*, *two* is not *duality*, *three* is not *threeness*, etc. Universals, furthermore, are Janus-faced entities. On the one hand they are expressed as properties of things—squareness belongs to objects—and, on the other, they are expressed as concepts—squareness is a concept that applies to things that are square. This Janus-faced quality mirrors Aristotle's use of the notion of *form*, for forms exist both in things and in the intellect. It is with the latter—universals as concepts—that we are concerned. It might be illuminating to the case under dispute if we could see how a naturalist might try to cope with the relation of intellect to universals, under the guise of concepts. We will do this with the following thought in mind. If naturalism cannot cope with the mind's grasp on abstract objects, in

the form of universals, which it must apprehend in order to think concepts, then perhaps we should not worry ourselves about the fact that naturalism cannot cope with the mind's apprehension of abstract objects, in the form of numbers conceived of as particulars: if naturalism cannot cope with thought at all, its failure to cope with arithmetical thought should cause no extra distress, and it would be wasted effort to try to devise a non-realist account of numbers to placate the already devastated naturalist.

1.2 Naturalism and grasping concepts

The naturalist tradition has two approaches to the mind's apprehension of concepts. Both are, in some sense, externalist, but one tries to allow for the conscious nature of conceptual activity and the other brackets off that issue. By 'externalist' in this context I mean any theory which denies that a mental or internal episode has any meaning-content intrinsically, and affirms that its content consists in some external relation to other things. The first expression of this was associationist psychology, and this theory does seem to make a gesture towards the conscious nature of concept apprehension, for the association in question is between the contents of conscious episodes. The point of this theory for our present concerns is that, in the concrete or empirical reality of the mental event of grasping the abstract object, what occurs is some purely particular occurrence which does duty for, or as, an abstract object—"redness," say—in virtue of its association with other particulars of the same kind. At its simplest, associationism is the theory that a mental content has the meaning it does because of the things that tend to follow it, or precede it, in the mind. So the word 'red' means red because thought of the word tends to be succeeded by a mental image of the color, or *vice versa*.

Associationism can be expressed in Hume's terms of the *tendency of the mind to pass* from one to the other, but one must be careful how one interprets such a tendency. It is tempting to think of it as a *felt* tendency, so that one is somehow aware of where the mind is going and so to apprehend the intension expressed by the association. This will not do, for putting it this way would be no different from saying that when one experiences the word 'red' one thinks or conceives of the color, whereas the point of the theory was to explain what such thinking, or

conceiving, is. It is similarly tempting to think of the color image as hovering in the background, somehow anticipated by the mind, when we entertain the verbal image "red." But such talk can only be metaphorical, for the image either is or is not in consciousness. And, contrary to Hume, being *very dimly* in consciousness is not the same as being thought of. The tendency of the mind to move from one thing to another has to consist in the straightforward fact that one thing usually follows, or is caused by, the other; the tendency or association cannot be thought of as some *experienced* feature of the situation without falling back to treating as primitive the mind's understanding of the intension that holds together the associated elements.

This is a point that needs emphasizing, because it is natural to interpret associationism as involving the subject's mind associating various mental images with each other. But such positive acts of associating would involve *recognizing the similarity* of the things associated with each other and hence would presuppose possession of the concept under which they are associated; such a theory would not be an analysis of concept possession. If one wants a completely naturalistic account of thought, the meaning of a sign or image must consist in the bare fact that it stands in an external causal relation to that which we say it signifies. From the point of view of the experiencing subject the meaningfulness is not something of which he is conscious; all he experiences is, first the word 'red,' then a mental image: there is nothing that could count as his internally associating them which does not reintroduce the mysterious generality of thought. It follows that associationism is a self-defeating theory, because, as I have just argued, the external relation that constitutes the meaning of the mental content is not itself something that the subject himself can apprehend: it is only constructible from a third-personal perspective. The reality of meaning would, therefore, only be accessible to some idealized third person who was able to observe the relation between the particulars. But there never could be such an observer—at least not if his thought-processes were to be analyzed in the same way as ours—because his thoughts about the relations of the particulars would themselves be just a succession of particulars whose relations, which give them meaning, were not directly accessible to him.

It will no doubt have struck you at the beginning of the discussion of associationism that, though this theory is part of a reductive

empiricist approach to thought, it is hardly a form of naturalism, because it takes phenomenal consciousness as basic, and phenomenal consciousness is something in the face of which, the naturalist can only, in C.K. Ogden's phrase, repeated by Ayer, "pretend that we are anaesthetized" (Ayer 1976, 125). The modern naturalist's approach to concept apprehension, therefore, tends to ignore consciousness, and, hence, takes a different route from the associationist.

Associationism is now something of a historical curiosity. The contemporary path for naturalizing conceptual activity is to follow a similar causal path to that employed for reference. Concepts are treated as representations and some formula of the following kind is endorsed:

A mental representation *R* is a representation of some feature *F* if *R* stands in some appropriately differentially sensitive causal relation to instances of *F*.

Such theories deliberately make no reference to consciousness. This itself is a serious shortcoming if you believe that the association between thinking and consciousness is not a mere accident. Even if one were to allow that a creature might standardly *think* without being conscious, would one wish to allow that it could *understand what it was doing* without being conscious? The notion of *understanding*, in this sense, has a reflective element that seems to involve consciousness. Someone who thought that a robot could understand what it was doing would need to adopt a behavioristic account of consciousness. The distinction between *performing intelligently* and *understanding what one is doing* obviously plays a major role in Searle's (1980) "Chinese Room" argument. But, putting aside these fundamental difficulties, within their own terms, the main problem for causal theories concerns the normativeness of conceptual content, and the absence of any such idea from causal relations. I cannot here go into this discussion in detail, but the situation seems to me to be as follows.

If one begins with a grasp on the intension of a concept, one can then devise a causal plan of the kind of mechanism that might realize or model the exercise of that concept, at least in its paradigm cases. Because one knows what concept one is trying to model, one knows which deviant responses of the system to ignore, rather in the way that one can draw a graph line through some dots, without each dot

necessarily falling exactly on the line. But such a procedure is plainly not an analysis of concept possession because it presupposes what one is hoping to analyze. One cannot construct the grasp on an intension “bottom up.” It is this fact that leads Kripke in his discussion of rule-following in the direction of a skeptical Goodman-like conventionalism (Kripke 1982). He sees this as the only alternative to “magic,” where ‘magic’ simply means treating the grasping of an intension, as opposed to an extension, as a basic idea. The skeptical conventionalism that Kripke attributes to Wittgenstein has been refuted more than once (e.g. Walker 1989, 142–145), and the need for a grasping of intension as a *sui generis* notion presupposed by any plausible account of thinking parallels the *sui generis* nature of phenomenal consciousness for any account of perception. The fact that naturalism can cope with neither makes it one of the great non-starters in the philosophical race.

The argument of this section can be summarized in the following form.

- (8) The notion of apprehending a concept/universal/intension is either basic and unanalyzable, or it can be analyzed in terms of relations between particulars.
- (9) If the latter, then either associationism or the causal theory of conceptual representation must be true.
- (10) Neither associationism nor the causal theory capture the notion of generality: rather they presuppose it.

therefore

- (11) Apprehending concepts (etc.) cannot be analyzed in terms of relations between particulars.

therefore

- (12) Apprehending concepts (etc.) is a basic, unanalyzable notion.

1.3 Misinterpreting the argument?

I have spoken as if Benacerraf’s argument sets up a conflict between naturalism and Platonism, but that is not how he puts it. Rather, the conflict is between “a reasonable epistemology” and Platonism. “A reasonable epistemology” shares with naturalism a commitment to

a causal component in the knowledge of particulars, but it does not necessarily carry with it a naturalistic account of concept possession. Someone who took the notion of grasping an intension as primitive might still insist that one is acquainted with empirical universals such as squareness by being in perfectly normal causal contact with square things. The causal element is no longer purporting to provide an analysis of what it is to grasp universals—this notion remains primitive: it merely provides a necessary component in one stage in the process, at least in the case of an embodied mind. Benacerraf *might* think this a perfectly reasonable epistemology, and he might think this even if he objects to it on other grounds.

The role of causality in an inference of the form “there are several square things, therefore there is a property of squareness” is unclear, and yet this appears to be a rational argument from observed facts to the existence of an abstract object. One does not need to be in causal contact with “squareness,” as opposed to an instance of it, to make this inference. A whole set of *a priori* relations—that it is a shape, all its Euclidean properties—can be inferred, Meno-like, from an instance. Is this very different from noticing groups of objects that instantiate numerical properties—for example, there being three objects in one group, two in another—and from the instances, abstracting the universal and it's *a priori* relations to others? Individuals can be known by description. “Squareness” could be known as “the property possessed by the outline of each surface of that die.” Frege, it seems, thought you could do something similar for numbers. Zero is explicated as the number that designates sets of things that belong to the concept *not self-identical*, and Hume's Principle carries you on to the other numbers. This could be thought of either as a system of definitions, using definite descriptions, or simply as a means of identifying the numbers by definite descriptions.

My point here is more negative than positive. It is that once one abandons the more reductive naturalist programme of explaining what the apprehension of the generality of a concept is, it becomes correspondingly unclear how much license one can allow oneself to work within the framework of the *a priori* starting from some causally based acquaintance with features of the world. If in any sense experience puts one in touch with universals, *a priori* reflection can acquaint one with the whole Platonic world.

There is another point. Once one allows that causal interaction with an instance of squareness can enable the mind to grasp squareness as such, in some irreducible sense, has one not already abandoned naturalism so strikingly, that one might as well allow that the mind can range “naturally” through the Platonic world?

1.4 The ontology of number and the Platonic realm

My positive contribution to this discussion consists in something that will seem to the comfortably modern to be a bizarre suggestion. This is that one can combine important features of fictionalism, Quineanism and Platonism by adopting what I will call, in a loose non-scholarly sense, a neo-Platonic account of abstract objects. I shall also suggest that there are good reasons for taking this path. For my purposes, the essential features of the three approaches are:

- (a) *Platonism*: numbers, and abstract objects in general, are *ante rem*; their existence is presupposed by and independent of the existence of individual objects.
- (b) *Quineanism*: numbers, and abstract objects in general, are legitimately thought to exist on the grounds that there could be no intelligible—“scientific”—account of the world without them.
- (c) *Fictionalism*: numbers—that is, numerical concepts or terms—do not correspond to any reality which is independent of the role they play in making the world intelligible.

These three features can be combined only if (i) the existence of an intelligible world presupposes the existence of a realm of abstract entities (which include numbers, but which also include some or all of universals, possibilities, propositions and maybe more): (ii) these abstract entities are better thought of as modes of understanding exercised by an objective intellect, rather than as “objects” in their own right.

How this approach accommodates Platonism is clear enough, because it starts from the acknowledgement that an empirical world requires *abstracta*. It is Quinean because it is in the context of an “enterprise” of rendering intelligible—that is, in relation to the activity of an understanding mind—that one needs to invoke these entities. And it is fictionalist because this ontology corresponds to nothing outside of

its role in rendering the world intelligible—though the world could not exist except as being intelligible in this way. The feelings that there is “something in” each of these approaches can be reconciled by making a necessarily existent objective *Nous* the home of the *ante rem* concepts, the thinker whose understanding requires that there be such objects, and who is the origin of their existence.

The line of thought which leads to this neo-Platonic conclusion is as follows.

- (13) Universals exist either as *in re* properties of things, or as *ante rem* self-subsistent entities, or as acts of the human mind (“conceptualism”), or as acts of some non-human mind—a neo-Platonic divine *Nous*.
- (14) Aristotle’s attack on Plato’s theory of Forms and Frege’s argument that concepts are not objects show that they cannot be both *ante rem* and self-subsistent entities.

therefore

- (15) Universals exist either as properties of things or as acts of minds, human or divine.
- (16) The need to provide an adequately realist account of possibility shows the inadequacy of both a purely *in re* theory of universals and of a purely conceptualist theory of a variety which confines them to human or finite minds.

therefore

- (17) Universals exist as acts of some non-human divine mind.

The argument is valid. The premises are (13), (14) and (16). If what was argued earlier is correct, and reductionist accounts of universality in thought invariably fail, then at least the third disjunct in (13)—conceptualism—must be true, though not to the exclusion of the others. The other disjuncts represent all the forms of realism. That some form of realism about universals is correct, I believe to have been convincingly argued by David Armstrong (1978). My own arguments for (16) below will strengthen this conclusion. Most of what follows will concern the arguments for (14) and (16).

2 The argument for neo-Platonism

2.1 *The defense of premise (14); Aristotle's criticism of Plato's theory of Forms, and Frege's concept horse*

It is my view that Aristotle rejects *ante rem* realism because he thinks that it involves self-predication and that this latter doctrine makes Forms an incoherent fusion between universal and particular. If this is to be a serious criticism of *ante rem* realism, and not just an *ad hominem* criticism of Plato's form of it, self-predication would have to be an essential component of any *ante rem* realism that treats abstract entities as self-standing objects.

This is not, I think, the normal approach to *abstracta* and their relation to self-predication. Self-predication is usually looked at as an excess, not a foundational feature. Most philosophers seem to believe that self-predication arises from Plato's confusion of the roles of forms as universals and forms as paradigms, the latter being needed only for the edifying dimension of his philosophy. My suggestion is that, in the minds of both Plato and Aristotle, self-predication is essential to the separate existence of forms—it is, in other words, in their view, an essential concomitant of an *ante rem* Platonic realism. Neither of them considers the possibility of what one might call “modern Platonism” which would allow *F*-ness separate existence as an abstract entity, of which it would be (for most *F*) a category mistake to predicate *F*. My substantive claim in this section is that the reason why neither Plato nor Aristotle consider the modern option is the same as—or very similar to—Frege's reason for thinking that “the concept *horse* is not a concept”; namely that an abstract *F*-ness would be essentially incomplete and so not an object of any kind (Frege 1980). Self-predication, on the other hand, seems to complete, or “saturate” the abstract entity, by making it not just a property, which is *of* something else, but also an instance of that property, which is a thing in its own right, something suited to be an object in the Fregean sense. Self-predication is, on this understanding, what makes possible Platonic forms as separately existing things. When he attacks Platonic “Ideas,” Aristotle is attacking the theory that universals are entities, which are separately existing *and* self-predicating, because these two features are necessary concomitants.

In order to make this clearer, I shall look at (1) what self-predication is and what its rationale is; (2) why Aristotle objects to the kind of separateness that is the dual of self-predication; (3) how this relates to Frege's problem with the concept *horse*; (4) how Aristotle develops what one might call a realist conceptualism which prepares the way for a neo-Platonic theory.

2.1.1 Self-predication and its rationale. It might seem quite clear what self-predication is: it is the doctrine that every form instantiates the property of which it is the form or universal; so the form of beauty is itself beautiful, etc. Taken in its natural sense, this is thought not only to be obviously false, but also to show Plato allowing what we might call his philosophical logic getting distorted by his high-minded metaphysics. The need to avoid this has led Fine (1993) to distinguish two interpretations of the doctrine of self-predication. One she calls *narrow self-predication* and this is the natural interpretation of the doctrine found above, according to which the form *F* is *F* in the most literal and simple sense. By contrast there is *broad self-predication*. On this theory it is sufficient to call the form of *F* '*F*' that it *explains* why other things are *F*. Also forms give the standard for being *F*—they are a kind of blueprint. I do not think that this will do, partly for the reasons I shall give below, and partly because I do not think that this account would explain why Plato thinks forms of this kind would do the “edifying” job that he does, indeed, think they can perform. Why should one abandon contemplating genuinely beautiful particulars in favor of contemplating something which merely tells you what it would be for something to be beautiful and which explains why other things are beautiful? An explanation, a standard or a blueprint exist for the sake of the things they explain or set the standard for, not *vice versa*.

Fine is, one can assume, motivated by a desire to save Plato from believing something manifestly false and rather silly. (Despite the fact that the “silly” doctrine is what, she believes, contemporaries like Aristotle attributed to him. It is a bad sign when a modern writer wants to make an ancient say what seem sensible to us, rather than what his contemporaries attributed to him.) I now want to show that, though what he said is, we can now see, manifestly false, it is not silly, because it was well motivated.

If Fine is wrong, does it follow that Plato is simply confusing philosophical analysis with edification? I think not, because the doctrine of

self-predication follows naturally, if not strictly necessarily, from some of the ways Plato argues for the existence of forms. It also provides a not initially wholly implausible answer to an apparent puzzle that we still have not answered, namely, that concerning what the intrinsic nature is of any given universal.

Plato's discovery was that the world cannot just be a collection of physical or spatio-temporal particulars, as the materialistically minded of the pre-Socratic thought. The general frame of the Platonic thought can be expressed as follows.

(18) We know that there are, in some sense, what we would now call "general concepts."

(19) These cannot be identical with spatio-temporal particulars.

because

(20) All particulars exhibit "compromised" versions of that which the general concepts express; any *F*-thing is also not-*F*.

therefore

(21) General concepts are entities of some radically different kind.

The crucial step in this argument is line (20). Three kinds of arguments can be used to defend this move, and Plato uses at least two. First, it can be argued that *F*-things are also not-*F*. This, in itself, can take at least two forms. It can either be a case of simple imperfection—nothing empirical is perfectly beautiful. This argument tends to be applied to normative concepts. Or it can be a matter of comparatives—something large in one context is small in another. Second, Plato argues that a particular thing that is *F* can *seem* not to be *F*, whereas the concept *F* itself cannot seem to be something else. Third, it could be argued that a particular thing cannot be the general concept because it possesses all sorts of other properties—anything large will also be square or heavy. Plato does not seem to use the last argument explicitly, though it is the most straightforward and universal one. This may be because it does not also support the conclusion that the empirical world is "less real" than the world of forms and so an appropriate object for opinion, rather than knowledge. But he does not seem to distinguish clearly between the ways an *F* thing may be

also not F —that is, by actually being the opposite of F , or just by being some other G .

This leaves the question of *what* kind of thing general concepts are—what is the intrinsic nature of these non-particulars? Line (20) above suggests an answer. They are what you would get if you could take a particular F and remove from it everything that is non- F —in the sense of all G that is not F , not just in the sense of all *opposites* of F . How could the product of such a process fail to be what F is in itself, for there is nothing else for it to be? The result is the self-predicating form. It is both concrete, in being the most perfect *instance* of F , and abstract or universal because it is the epitome of that which all F -things share.

2.1.2 *Aristotle's objections to this conception of form.* Aristotle's criticism of Plato's theory is that one cannot have an entity that is both an individual and a universal. He agrees with the Platonists that *if* forms are separate entities then they must be self-predicating.

They make this mistake [of treating universals as self-predicating substances] because they cannot say which substances are thus indestructible ... So they make them the same in form as the destructible ones we know, simply adding the word 'itself' to perceptible substances, as in 'man-itself' and 'horse-itself.' (*Met.* 1040b29–34, translation from Bostock 1994, 29)

In other words, the status of forms as indestructible substances requires that they be endowed with some positive nature, and hypostasizing the feature as it occurs in the world is the only thing that falls to hand. This makes it clear that Aristotle would have no sympathy with “modern Platonism”: if there really were abstract entities existing in their own right, they would have to be self-predicating in order to give them a nature. But such bastard conflation between individual and universal, according to Aristotle, give rise to contradictions. The simplest of these is that universals can be defined and individuals cannot (1040a8ff.). Another is that an individual cannot be in many different locations at once, but a universal can (1039a24ff.).

Do these arguments really work against a modern platonism? Here we turn to Frege.

2.1.3 *Frege and the concept horse.* Frege notoriously said that the concept *horse* is not a concept. It is not my intention to enter deeply

into Fregean exegesis, but rather to draw attention to and make use of the parallel between Frege's and Aristotle's objections to the objectification of universals. Frege seems to maintain four doctrines.

- (22) Concepts are unsaturated and cannot exist on their own.
- (23) Objects are saturated and can exist on their own.
- (24) Predicates refer only to concepts.
- (25) Names—which include definite descriptions—refer only to objects.

From (22) and (23) it follows by the law of identity that

- (26) Concepts and objects are different things.

and from (25) and (26) it follows that

- (27) Names cannot refer to concepts.

As 'the concept *horse*' is a name, it must refer to an object, so what it refers to—i.e. the concept *horse*—is not a concept. From a Fregean perspective, the least radical way of avoiding this paradox would be to deny (25) in its generality. One could argue that it is true of first order names, but that one can have second order names that refer to concepts. The issue is then whether something can stand in the object *position* in a proposition without actually being an object, or whether this is a kind of ill-formedness. The more radical rejection of the paradox, and one which would be essentially un-Fregean, would be to deny (22). My concern is to draw attention to the similarity between Frege's doctrine that concepts are unsaturated and cannot stand alone to Aristotle's objection to Platonic forms. Using Frege's language, Aristotle's objection is that Platonic forms are objects and, as such, cannot serve the function of concepts, which they are intended to do. The price of treating them as separate entities is to give them logical features that make them unsuitable to be multiply instantiated. This suggests that Dummett at least oversimplifies when he says

the fact is that the notion of an "object" itself, that is, the notion as used in philosophical contexts, is a modern notion, one first introduced by Frege ... According to the ancient tradition, entities are to be characterized as particulars and universals. It is characteristic of particulars that we can only refer to them and

predicate other things (universals) of them. (Dummett 1973, 471)

In other words, according to Dummett, before Frege there were only different kinds of *objects*, the radical distinction between object and concept was not appreciated. If my interpretation of Aristotle is correct, it was appreciated at least by him, to the extent that he saw that universals had to be incomplete entities if they were to fulfill the role of applying to many things, and that making them into just another kind of object made it impossible for them to fill this role.

One modern response to the radical distinction between objects and concepts has been to see it as licensing the abandonment of the quest to understand the ontological status of universals: only objects enter into ontology so concepts and predicates need no ontological grounding in their own right. The thought contrary to this is that in the truth conditions of sentences such as 'Fred is a man' and 'Fred is fat' there is something that answers to 'is a man' and 'is fat' just as there is something that answers to 'Fred,' and it is the nature and status of such things in which we are interested in the problem of universals. I shall assume that the second thought is correct and that the problem of universals is real. Granting this, where do Aristotle's and Frege's criticisms of Platonism leave us, for they rule out one—in a sense, the simplest—of the possible answers to the problem, namely that universals name another set of "things," which stand alongside the "things" to which names refer? The Fregean suggestion that second order quantifications such as 'the concept *horse*' introduce another set of objects that stand in for the concepts so that we can refer to them indirectly seems hopelessly contrived and, more important, not to answer the issue of *what it is* that we are indirectly referring to. These new objects cannot be the real ontology doing duty for something we need (concepts) but which, because of their logical type, cannot actually affirm directly to exist.¹

¹ It has been pointed out to me that Frege at certain times considered the possibility that expressions such as 'the concept *horse*' refer to the extension of the concept. (See Burge 1984.) Such a proposal would still not solve the question of *what a concept is*. Frege is some kind of Platonist—he does not think that *concepts* are their extensions. We still need an answer that does not turn them into objects.

2.1.4 *Aristotle's theory of form and universal.* What is Aristotle's own theory? He can be read as a conceptualist. On this reading, the forms in things are particulars—so-called “individualized forms”—and universals are a creation of the mind. They are the “forms without matter” that constitute thought. There are plenty of texts which support both these points. On the other hand, forms in the intellect are not simply creations of our minds. The active intellect contains all forms, and it is tempting to claim that the prime mover, in thinking about itself, thinks all forms too. There are very famous and probably hopeless mysteries of interpretation here. But there is a philosophical argument that can help, if not with the exegesis of Aristotle, with the philosophical development of the resources he provides.

The doctrine of individualized forms, when combined with a conceptualist, *post rem* theory of universals, is rather like the theory nowadays called *tropism*. It implies that, in the absence of human thinkers, or when some property is still undiscovered, there could be tropes or individual forms of *F* and no corresponding universal. This seems to me like saying that there could be an *instance of F* without there being *F per se*. But I do not see how the notion of an instance of a property could be more primitive than that of the property itself. To avoid this objection the tropist-conceptualist would have to find some adequate way of characterizing the trope that did not make mention of its relation to the property. Given that particular qualities are still qualities, this is equivalent to saying that the hyphenated *instance-of-F* is primitive, not defined in relation to *F* itself. There seem to me to be two general kinds of accounts concerning how one gets from this primitive to the universal, and understands the relationship between them. One I shall call “modern,” the other “Aristotelian.” According to the modern theory, *F* itself is constructed from the similarity between instances-of-*F*. As far as I can see, this move will not do, unless it goes so far as to collapse into the resemblance theory of universals. The tropist believes that the similarity between *F*s is grounded in the “particular” nature or quality of the object: this is what distinguishes him from the resemblance theorist, who thinks resemblance is prior to any quality or nature. But can one treat a *quality* as wholly particular—is it not endemic to the idea of a quality that it can have many instances? The tropist may say that this begs the question: tropes are particulars, which may resemble each other even to the point of exact resemblance. But such a theory involves taking different facets or aspects of objects as

basic, for the particular shape or the particular color are aspects of the particular object. And the question arises what these aspects—shape or color—are if not generic universals. When comparing objects for shape (for example) one can pick out those aspects one should compare—one does not end up comparing by mistake the shape of *a* with the color of *b*. If these aspects are not to be reified into *kinds*—a universalist notion—then they would have to be further analyzed into similarity classes for objects as a whole, and one is back to pure resemblance theory.

The Aristotelian theory is different because it does not seek to *construct* universality. One could truly attribute the following to Aristotle: the form *F* is something which is particular when embodied in a (physical) particular and universal when thought. Universality is a product of thought, but only because when form is made an object of thought it is thereby rendered universal. But one cannot give all the credit to the nature of thought. It is not thought alone that creates the universality, but also the nature of form. When one thinks that Fred is fat, merely making Fred an object of thought does not turn him into a universal. Names and referring expressions remain essentially different from properties even when thought. Form without matter is naturally universal. But then what are we to say about that which binds or is common to the individual form in the object and the form-without-matter in the mind? This cannot be analyzed in terms of resemblance, for a thought concept and an instance of a property cannot be said literally to resemble each other. Some primitive notion of being *the same in account* or in *logos* is unavoidable. The universality immanent in the particularized form is not eliminable. Nevertheless, it is possible to interpret Aristotle as saying that immanent is all that it is. For the modern tropist, universality is a construct from abstract particulars and similarities. For Aristotle, his “abstract particular”—the individual form—is naturally universal *when freed from matter*: the universal is not somehow a logical construction, but a property of the form considered neat.

If we cannot manage with universality as a wholly *post rem* creation, and if Platonism is unacceptable, this brings us naturally to *in re* realism, and to consider it as an interpretation of Aristotle. *In re* realism respects the requirement that universals be incomplete objects, because, according to this theory, they exist only as properties of objects, and not in their own right. Before setting the ground for the proper discussion of *in re* realism, which I can only do briefly here, I want to make one important

observation about its relation to a belief in individualized forms or property instances. Many realists—Platonic or *in re*—claim that if one believes in universals and particulars, one does not also require *instances*. Contrary to this, it seems to me that every realist believes in instances, whether they realize it or not. Armstrong, who denies instances, believes in the *state of affairs* of *a*'s being *F*, in addition to *a* and to *F*. A Platonist, who only believes in the form and the particular cannot fail to distinguish the fact or occasion of *a*'s being *F* from the fact or occasion of *b*'s being *F*. In other words, everyone needs some third element in their ontology that picks out and differentiates the particular coincidences of form or universal and object. If one tends towards an ontology of substances and their properties, one will express this in terms of *instances* of the universal. If one tends to see the world as the “totality of facts, not things,” one will pick on facts or states of affairs. Either way, one cannot avoid something over and above the universal and particular which captures the being of that universal in that particular. This is mainly what talk of individualized forms does. This explains why it can be so difficult to tell from the text whether Aristotle is an *in re* theorist or not: whatever his theory, it is virtually impossible to avoid talking in both ways: with reference, that is, both to *what is common*, which is like talking about the universal in the thing, and to the *causally efficacious presence in a given case*, which is the instance, or the fact or state of affairs of its being present.

We can see, therefore, that Aristotle's theory is a mixture of conceptualism and realism. Only in a mind does a universal manifest itself completely, but the individualized form in a particular object is not a straightforward particular, being immanently universal. And the universal in the mind is not simply a *post rem* created concept, but something which, like the active intellect itself, is also eternal.

2.2 The defense of premise (16)

2.2.1 *Introductory remarks.* (16) denies two of the options in (13), and denies them both on the same ground, namely that they are inconsistent with realism about possibility. The two positions that it denies are (a) an *in re* realism that denies that there are any uninstantiated universals, and (b) a conceptualism that restricts universals to concepts formed by the human mind. The arguments against both (a) and (b) share the same first premise and have parallel structures, as follows.

- (28) If one is a realist about unactualized possibilities—that is, if one thinks that there are objective facts about what is and is not possible—then one must be as realist about the universals involved in those possibilities as one is about actualized universals.

therefore

- (29) If the *in re* theory is true and there are no uninstantiated universals, then the only real possibilities will be those involving actualized universals: that is, the Combinatorial Theory of Possibility will be correct.²
- (30) The combinatorial theory is not correct.

therefore

- (31) If realism about unactualized possibilities is correct, then there must be uninstantiated universals.

But it also follows from (28) that

- (32) If conceptualism is correct about universals, then there will be no possibilities involving properties that have not been conceived by the human mind.
- (33) The restriction of possibilities in this way is incorrect.

therefore

- (34) If realism about unactualized possibilities is correct, then there must be universals not conceived by the human mind.
- (35) Realism about unactualized possibilities is correct.

therefore

- (36) Neither the *in re* theory nor conceptualism is correct; that is, (14) in the original argument is correct.

² It has been pointed out to me by the editors that the *in re* theory is compatible with the existence of uninstantiated or alien universals, because it is compatible with Lewis's version of modal realism. On a trivial level, the debate might be represented as verbal, because, for Lewis, all possible universals are instantiated in some world or worlds. Nevertheless, the point is a substantial one: an *in re* theorist could consistently be a Lewisian and not a combinatorialist, in the ordinary understanding of that term. A full discussion of the *in re* theory's ability to cope with possibility, therefore, would have to include an evaluation of Lewis's theory. In this context I shall only say that I regard the usual reasons for rejecting that theory to be satisfactory.

The premises that require defense are (28), (30), (33) and (35).

(33) and (35) go naturally together, because they both claim that there are real object truths about what might happen, or might have happened, which do not depend on the human mind. This must surely be so if there is randomness in the physical world, for there cannot be randomness unless it is really possible that any of a variety of things might occur.

The rationale for (28) is that, if there exist facts which can be expressed in the form “something might have been *F*” or “something might (yet) be *F*,” and such facts obtain independently of any human mind, then there must be the things necessary for the constitutions of such facts, which, in this case, includes the universal *F*.

Allowing that these rather brief arguments cope, for present purposes, with (28), (33) and (35), the crucial premise is (30)—the denial of the Combinatorial Theory of Possibility. I have discussed this at length elsewhere (Robinson 1998), and here I restate the intuitive argument for why the range of possibilities should not be constrained in the way CTP requires.

2.2.2 The refutation of the Combinatorial Theory of Possibility and of conceptualism. The Combinatorial Theory of Possibility is not plausible, because it cannot accommodate the logical possibility of what might be called *creativity at source*. This problem first struck me in the form of a disproof of the existence of God, which follows from combinatorialism. Whatever one thinks of the conclusion of the argument, as an argument it must strike one as fishy, though clearly valid.

(37) If there is a God then He could have created a world with at least one simple property different from those that actually exist.

(38) If something could have been created, then that thing is possible.

therefore

(39) If there is a God then it is possible that there should have been a world with at least one simple property different from those in the actual world.

(40) If the Combinatorial Theory of Possibility is correct, there could not have been a world with any different simple properties from those that actually exist.

therefore

(41) If the combinatorial theory is correct then God does not exist.

(42) The combinatorial theory is correct.

therefore

(43) God does not exist.

But also

(44) If the combinatorial theory is correct it is necessarily correct. (The correct account of modality cannot be only contingently true.)

(45) Anything that follows from a necessary truth is necessary.

therefore

(46) Necessarily, God does not exist.

So God is impossible, according to combinatorialism. The source of the problem is that combinatorialism does not allow for the potentiality for a simple property that is not actual; only combinations can be potential but not actual. A similar counterintuitive conclusion follows without invoking anything as contentious as God.

(i) Let us suppose (as seems plausible) that at least some simple properties were created randomly in the first nano-seconds of the "big bang."

(ii) If combinatorialism is correct, then two unacceptable consequences follow. (a) Though it is possible that some or all of those randomly created properties might not have been created, it is not possible that different ones might have been randomly created. (b) Nor, supposing that the random production of the actually produced property had not taken place, would it have been a possible property, or a possibility that it might have been produced.

In so far as what we have here are random productions of simple properties, this seems totally unreasonable.

This latter case, (b), is one that Armstrong (1989) considers and bites the bullet: if there had been a sparser world than this one is, then this world would not have been a possible world. Now it seems at first sight that anything that has the consequence that our actual world might not have been possible, in a broad sense of possible, is plainly

false, for what is actual could not fail to be possible. Armstrong there took a hard line.

Given our stand on alien universals, symmetry of accessibility then fails. For a set of worlds which contains both [the actual and a contracted world] one must content oneself with an S4 modal logic. (Accessibility becomes reflexive and transitive, but not symmetrical.) (62)

With actuality contracted, possible worlds must be considered contracted. (63)

My problem with this is that the very idea of restricted accessibility rests on the idea that, amongst all possible worlds, some are accessible and some are not: there is, that is, a contrast between broad and narrow possibility built into the idea, and Armstrong has an account of only the narrow one. If there are inaccessible possible worlds, then they are, in an important way, possible.

Armstrong wanted to deal with these broader possibilities as what he calls “conceivable” or “doxastically possible,” rather than possible, proper. A merely conceivable world is one formed combinatorially from *what seem to us to be* independent simple properties, when in fact they are either confused perceptions of complexes that seem to be but are not simple (as with color concepts), or they are conceptions of things that are impossible, as with unrealized simples. Armstrong seems to talk as if the difference between real possibilities and mere conceivabilities was like that between the real and the imaginary. In fact they are just two kinds of fiction, one following slightly tighter rules than the other. I cannot see what the question of what is *really* possible under these circumstances comes to. In particular, if I say that it is conceivable, but not possible, that the “big bang” might have thrown up some different simple properties, what am I getting at? When the conceivable but not possible is not a function of our ignorance or confusion, it is not clear what work the distinction is doing.

Actually, I think the combinatorialist does have a way of coping with God, but the consequences are illuminating. Supposing Armstrong’s philosophy of mind allows for the possibility of a God—presumably as immaterially realized dispositions of some kind—then it

seems to me that everything that that being could have created would have been possible. How not, if there is a being capable of realizing them? So Armstrong ought at least to extend his theory to include all simple properties that are conceived by a mind that has the power to realize them if it should so choose. Most of us, it is plausible to argue, can perhaps conceive of but cannot realize new simple ideas, but God could realize those He does not choose to, as well as those He does. Armstrong is worried by the lack of truthmakers for possibility statements related to unreal simples, but if they and their potentialities are conceived of by someone with the power to realize them, then there are truthmakers, namely states of the relevant mind. Notice that the truthmaker problem still exists for the random property in the "big bang" case, unless there is a God as well as a bang.

The defense of the argument for neo-Platonism has now been completed, but the idea that the abstract Platonic realm is not a collection of abstract objects, but divine mind will strike almost all readers as so bizarre that I need to say something in the vain hope of making it less unacceptable. This is the task of section 3.

3 Why the neo-Platonic account of abstract objects and "what makes the world intelligible" is not as strange as it might seem

Treating the Platonic realm as an intellect and not as a collection of inert abstract entities enables one to preserve the virtues of a variety of theories about the world of *abstracta*. In an obvious way, it preserves the essence of the *ante rem* theory. At the same time, it preserves Aristotle's conceptualism, for what are abstract *objects* in conventional Platonism become *concepts* in—constituents of the thoughts of—*Nous*. The Quinean view of *abstracta* as tools for understanding is also preserved, though without the priority granted to purely human understanding. This latter is, of course, a major difference, but it preserves the important insight that these things are the product of the interaction of intellect and world, with the additional thought that, without intellect, there could be no world. The situation with fictionalism is similar. Fictionalism maintains that abstract objects are invented by the understanding minds as tools for making sense of the world. In the

case of fictionalism proper, this mind is the human mind and the world itself has no need for these contrivances, which is why they are fictions. On the neo-Platonic account, the mind in question is not ours, but objective *Nous*, and there could be no world that was not subject to its thinking. The abstract objects by which *Nous* thinks are creations of a mind, like fictions, but, unlike fictions, are essential to the constitution of the reality that is thought about.

I want, however, to end by broadening the role of intellect from that of thinking the usual range of abstract entities. It seems to me that there are many kinds of fact which are both essential to the reality of a world, and yet impossible without the presence of intellect.³

Philosophers are very familiar with the distinction between fact and value, and with the idea that the factual statements concern the world as it is in itself, but value judgments report the world in the light of certain human attitudes. (In saying that philosophers are familiar with this distinction, I am not implying that they all accept it, or that they should.) I want to make a parallel distinction within the realm of the factual between what I shall call *pure description* and *factual assessment*. The former represents the world as it supposedly is in its own right, the latter presents a kind of fact that is present only under the gaze of an assessing mind. The important point, however, is that there can be no world without the latter.

There are many true and important kinds of facts about the world which seem to be a function not only of how the world coldly and simply in its own right *is*, but to be a product of how the world is and of a judgment or assessment of how the world is. A simple illustration of this is as follows. The easiest kind of fact to take in the pure descriptive sense is one that concerns the location of a particular object in space and time. Thus the thought or proposition “there is a pen on the table,” if true, simply reflects a situation in the world. It is made directly true by what there is in concrete reality. But negative propositions do not “picture reality” in the straightforward way that some positive assertions do. The proposition “there is no pen on the table” is not so simple as its positive counterpart. *Absences* are not part of concrete

³ The argument in this section is developed at somewhat greater length, see Robinson 2009.

reality. A fact of an absence is more like an assessment of a situation with a certain thought or perspective in mind—in this case, a concern about a pen or pens. It is *grounded* in the concrete situation—what the situation positively is on the table's surface—but it is not simply a representation of that alone. *If* one wishes to include negative facts in one's ontology, a simple physical realism does not seem to be available.

Even the category of *states of affairs* is more problematic than it may seem. David Armstrong (1997) believes that the world includes states of affairs and that everything in the world is spatio-temporal: hence that states of affairs are spatio-temporal. Indeed, anyone who wishes to include states of affairs within his physical ontology, realistically conceived, would seem to have to deem them spatio-temporal. But it is not clear that they have spatial location. Perhaps it is easy enough to assign location to the states of affairs of macroscopic objects' possessing the traditional primary and secondary qualities. Thus a ball's being round will be where the circumference of the ball is, and its being red, where its surface is. With more arcane properties, the issue is harder. A particular electron has location, but where is its having spin or its having mass? Are they everywhere the electron is? Relations also create further problems. One could draw a *rough* outline round the state of affairs of *a*'s being six inches away from *b*: it would include *a* and *b*, but what of the intervening space would it include? In Timothy Sprigge's words

The point is that there is no distinguishable portion or piece of reality which is where the relation is exemplified as there is in the case of a property ... There is not, so to speak, some sub-division of the totality of particular reality which actualizes the relation to the exclusion of its contraries as in the case of properties. (Sprigge 1983, 164)

A pure description is a characterization of the world the content of which is the same as the content of its truthmaker. In other words, the truthmaker can be identified with the state of affairs that is asserted to obtain in a pure description. A factual assessment is a characterization of the world whose content differs from the content of its truthmaker. It seems to me that at least certain kinds of factual assessments are true and essential to our notion of a real world. If this is so, such a world could not exist without minds. A possible example of this is the following.

Suppose a certain random event (for example, the decaying of a particle) occurred at t , but that it could have occurred at t' , and that had it occurred at t' , certain very different things stretching into the future would, or could, have happened. If you are a realist about possibilities, you will want to say that it is a fact that these things would, or could, have happened, whether or not there are any minds. The truthmakers for such statements will be certain immediate dispositions or powers of actual objects. The extrapolation to more remote future outcomes cannot be thought of as existing in these actual states of affairs. This is shown to be true by the case in which, had the event occurred at t' , not t , the future that ensued could itself have gone either of two ways depending on some further random event. Let us express what actually occurred by p . What might have occurred is q . If q had occurred, then it might have been followed either by r or by s . The 'might' here is not epistemological, but is based on genuine indeterminacy. So if p had not been true, but q had, then either q followed by r could have been the case, or q followed by s . The truthmakers for either of those possibilities in the actual world would be the same. This is so because if there is a circumstance C in which there are a set of outcomes that are indeterministically possible, it is the same set of facts that underlie the possibility or probability of any of those outcomes, because they determine the range of probabilities as a whole. Nevertheless, it could be argued that the realist and common-sense intuition is that it is objectively true that there are the two possible outcomes, whether or not there are any ordinary minds to appreciate this fact. If this is the case then factual assessments, and, hence, some form of mentality, seems to be presupposed by our common-sense conception of the world.⁴

⁴ The editors have provided the following response to this argument. One may claim that the truthmakers for the various possibilities consist in the probabilities derived from what has happened in similar cases. This is a very serious suggestion, which I do not have space to discuss properly here. My main reservation about it is that treating the possibility of something happening in a given case as being made true by what happens in other cases (in other words, these other cases are not merely the evidence but the truthmakers) is not a realist but a logical constructionist approach to the possibility in question. And logical constructions are factual assessments, not pure descriptions. I certainly do not pretend that these remarks close the matter.

Conclusions

There are three major conclusions. The first, defended in section 1, was that, as there is no acceptable naturalistic or reductive account of what it is to grasp universals and think them as concepts, there is no particular problem about our apprehension of numbers: both universals and numbers are abstract objects and there should be no essential disparity in the accounts of how our intellects can apprehend them. Benacerraf, if he should be worried at all, should be worried about how thought in general is possible. The second, defended in section 2, is that the fact that universals or concepts cannot be treated as objects, together with the inadequacy of the *in re* theory of universals, shows that the best way of thinking of them is as being the thoughts of a divine intellect, in the manner of the neo-Platonists. The third, defended in section 3, is that such an intellect is, on other grounds, necessary if we are to take the existence of the world realistically.⁵

References

- Armstrong, D.M. 1978. *Nominalism and Realism*. Vol. 1 of *Universals and Scientific Realism*. Cambridge: Cambridge University Press.
- . 1989. *A Combinatorial Theory of Possibility*. Cambridge: Cambridge University Press.
- . 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Ayer, A.J. 1976. *The Central Questions of Philosophy*. Harmondsworth: Penguin.
- Benacerraf, P. 1973. "Mathematical truth." *Journal of Philosophy* 70: 661–679.
- Bostock, D. 1994. *Aristotle: Metaphysics, Books H and Z*. Translation with a commentary. Oxford: Clarendon Press.
- Burge, T. 1984. "Frege on Extensions of Concepts." *Philosophical Review* 93: 3–34.
- Dummett, M. 1973. *Frege: Philosophy of Language*. London: Duckworth.
- Fine, G. 1993. *On Ideas: Aristotle's Criticism of Plato's Theory of Forms*. Oxford: Clarendon Press.

⁵ I am grateful for comments from other participants in the conference and particularly for careful criticisms from the editors.

- Frege, G. 1980. "On Concept and Object." In *Translations from the Writings of Gottlob Frege*, edited by P.T. Geach and M. Black, 42–55. Oxford: Blackwell.
- Kripke, S. 1982. *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.
- Robinson, H. 1998. "Problems with the Combinatorial Theory of Possibility." *Acta Analytica* 21: 147–161.
- . 2009. "Idealism" in *The Oxford Handbook of the Philosophy of Mind*, edited by B. McLaughlin. New York: Oxford University Press.
- Searle, J. 1980. "Minds, Brains and Programs." *Behavioural and Brain Sciences* 3: 417–457.
- Sprigge, T. 1983. *The Vindication of Absolute Idealism*. Edinburgh: Edinburgh University Press.
- Walker, R.C.S. 1989. *The Coherence Theory of Truth*. London: Routledge.

About the Authors

ZSOLT NOVÁK has recently obtained his Ph.D. from the University of Leiden and from Central European University for his thesis book *A Priori Truth in the Natural World: A Non-Referentialist Response to Benacerraf's Dilemma* (2010). He is the co-editor of *Ész, Természet, Történelem* (2002). His research interest falls within the confines of the philosophy of language, metaphysics and epistemology.

ANDRÁS SIMONYI is Research Fellow at the Applied Logic Laboratory, Budapest. His main areas of research interest are semantics, formal ontologies and philosophy of mathematics.

DANIEL ISAACSON is University Lecturer in the Philosophy of Mathematics at Oxford University and Fellow of Wolfson College, Oxford. He has held academic appointments at the University of Washington and the Rockefeller University, and a visiting appointment at the University of California at Berkeley.

NENAD MIŠČEVIĆ is presently Professor of Philosophy at the Philosophy Department of the University of Maribor, Slovenia and Recurrent Visiting Professor at Central European University, Budapest. He is the author of a series of papers on intuitions and of several books, two in English, one on rationality and cognition published in Toronto, and another by CEU Press.

IAN RUMFITT is Professor of Philosophy at Birkbeck College, University of London; he previously taught at the Universities of Oxford and Michigan (Ann Arbor). He has published on philosophical logic, the philosophy of language, and the history of analytic philosophy

(especially Frege). A monograph on the philosophy of logic is forthcoming from Oxford University Press.

TIMOTHY WILLIAMSON is Wykeham Professor of Logic at Oxford University. His books include *Identity and Discrimination* (1990), *Vagueness* (1994), *Knowledge and its Limits* (2000) and *The Philosophy of Philosophy* (2007). He is a Fellow of the British Academy and a Foreign Honorary Member of the American Academy of Arts and Sciences.

RALPH WEDGWOOD is a Professor of Philosophy at Oxford University and a Fellow of Merton College, Oxford. He is the author of *The Nature of Normativity* (2007), and of numerous articles on ethics (especially metaethics, but also normative ethics and the history of ethics), on epistemology, and on the theory of rational choice.

HOWARD ROBINSON is University Professor in the Department of Philosophy at Central European University, Honorary Visiting Professor at the University of York and an Honorary Fellow of the Department of Philosophy at Liverpool University. He has published mainly on the philosophy of mind, the philosophy of perception, idealism and the history of philosophy.

Index

- absolute generality, xxvii–xxx, 52–53, 192–203
- abstract objects/entities
- and Benacerraf's dilemma, 235–237, 261
 - fictions as, 37
 - grasping, xxxii, 235–237, 261. *See also* apprehension of universals/concepts/senses
 - inferring the existence of, 241
 - as mathematics's subject matter, x, xiii, xviii, xx–xxi
 - neo-Platonic account of, xxxiii, 242–243, 257–258, 261
 - numbers and sets as, 37
 - and self-predication, 244, 247
 - structures as, 20
 - See also* forms; mathematical objects; tropes; universals
- abstract particulars. *See* tropes
- abstraction, 10, 241, 27n10
- acceptance, attitude of, 148–152
- alien universals. *See* uninstantiated universals
- Anscombe, G.E.M., 216
- ante rem* structuralism, xviii–xix, xxi–xxii, 31–34, 37, 42
- ante rem* theory
- of abstract entities, 242, 257
 - of concepts, 243
 - of numbers, 242
 - of structures. *See ante rem* structuralism
 - of universals, 33, 243–244. *See also* Plato's theory of Forms; realism: about universals
 - See also* Platonism
- anti-realism
- and Benacerraf's dilemma, xvi–xvii, 101
 - its contrast with realism, xv–xvi
 - about logic, 82, 86–87, 100–101, 120
 - about logical concepts/senses, xxiii, 120
 - about mathematical referents, xviii
 - about truth, xvi–xvii, 101
- apprehension of universals/concepts/senses, xxxii–xxxiii, 235–242, 261
- of logical concepts/senses, 89–92, [93–94]n15, 102–103, 120
 - See also* grasp
- Aristotle, 33, 235–236, 243–245, 247–252
- arithmetic
- Benacerraf on, 19
 - Dedekind's axioms for, 24–25, 28, 36, 42, 56
 - explaining *a priori* knowledge of, 110
 - Frege on, 7
 - if-then-ism about, 40–41
 - non-standard models of, 28, 38, 67, 186n11
 - Peano (PA), 5, 9, 24, 48, 54, 67
 - second-order, 3n1, 40–41, 46, 186n11
 - undecidable problems in, and CH, 48, 66
 - See also* natural numbers; truth: arithmetical
- Armstrong, D.M., 243, 252, 255–257, 259
- assessment
- of beliefs and assertions, 229
 - factual, 258–260
 - of soundness/validity, 128–133, 137, 155–156, 195–196
 - of utterances and statements, 220–222
- associationism, 237–240

- Atomic Freedom, Rule of, [198–199]n24
- Axiom
- of Coherence, 28–29
 - of Foundation, 49–50
 - of Infinity
 - Shapiro's, 27–28
 - Zermelo's, 15, 37
 - of Replacement, 4, 11, 15–17, 45, 50, 54–55
 - Tarski's, of Inaccessible Cardinals, 51
 - $V=L$, 59
- axiomatic method, 24, 30, 146
- Ayer, A.J., xiv, 214–215, 217, 239
- Balaguer, M., xx
- Basic Law V, 7, 66
- Bealer, G., 212n, 229n21
- Belnap, N., xv
- Benacerraf, P., ix, xxi, 2, 18–20, 63, 67n, 241. *See also* Benacerraf's dilemma
- Benacerraf's dilemma, x–xi, xx–xxi, 2, 81, 101, 235–236
- and apprehension of universals, xxxii, 236–237, 240–241
 - for logic, xi, xxiv–xxv, xxvii, 77–78, 82, 102, 119–121
 - and obviousness and compellingness, 121
 - presuppositions of, xi–xiv
 - as a problem for rational insight theories, 88
 - responses to, x–xi, xiv–xx
 - about senses, 102, 120
- Bernays, P., 10, 12–13, 16–18, 20, 47. *See also* von Neumann–Bernays–Gödel set theory
- Blackburn, S., xv, xvii, 207, 211–213, 216n, 224
- Boghossian, P., 78, 88–90, 97n, 103n, 104n
- BonJour, L., xx, 82, 88, 106, 111
- Boolos, G., 46, 53, [66–67]n21, 68, 201n
- Borel Determinacy, 4–5, 53–54, 67
- Bourbaki, N., 22–25, 32
- Brouwer, L.E.J., xvii
- Brown, J., xx
- Burali-Forti paradox, 27, 202
- Burge, T., xxiv, 96, 249n
- Butković, A., 97n
- Camp, J.L., xv
- Cantor, G.F., 11–12, 30n, 47, 58, 61, 65
- continuum hypothesis. *See* continuum hypothesis
 - cardinals, 49–50, 58. *See also* inaccessible cardinals; large cardinals; Mahlo cardinals; Woodin cardinals
- Carnap, R., xvii, 12, 44, 62
- Casullo, A., 81
- categorical characterizations, 2–3, 34–35, 38–39
- of the continuum, 2, 5
 - of the cumulative hierarchy, 10, 15, 38. *See also* quasi-categoricity
 - of the Euclidean plane, 2–3, 7, 42
 - of the natural numbers, 2, 32, 36–38, 47
- categoricity. *See* categorical characterizations; quasi-categoricity
- causal accounts
- and conceptualism about logic, 78, 89–90, 107–108
 - of impressions of validity, 100
 - of logical beliefs/intuitions, xi, 122
 - of the meaning of representations, 238–240
 - of perceptual and of *a priori* knowledge, 81–82, 109–110. *See also* causal contact theory of knowledge acquisition
- causal contact theory of knowledge acquisition
- acquisition, x–xi, xiii, xviii–xix, xxxii, 81, 235–236, 240–241. *See also* causal accounts; contact/no-contact theory of knowledge acquisition
- CH. *See* continuum hypothesis
- characterizations. *See* categorical characterizations; unique characterization
- Cherniak–Stich program, 111
- Chisholm, R., 109
- circularity
- of categoricity proofs for set theory, 3, 34–36, 64–65
 - in justifications of logical principles, 82, 90n, 92, 97–98, 104n, 114, 120
 - of mathematical theories of structures, xxi, 27
- See also* non-circularity constraint; single-domain problem

- cognitivism/non-cognitivism, xii, xiv–xvi, 210–211. *See also* truth-conditional semantics
- Cohen, P., 12, 16–17, 53, 55, 57–59
- coherence
- of absolute identity, 191–192, 201
 - of beliefs, 98–99
 - of characterizations of structures, xxi–xxii, 28–29, 32, 41
 - and formal consistency, 28
 - of formulas, 28–29
 - of the notion of cumulative hierarchy, 14
 - of second-order Zermelo-Fraenkel set theory, 29
 - Shapiro's Axiom of, 28–29
 - of theories/accounts, 9, 28–29, 41
 - of unrestricted quantification, 197
- coherentism, 80, 98–99. *See also* coherence
- Colyvan, M., 87n
- Combinatorial Theory of Possibility, 253–257
- commands, 215, 221–222
- compellingness. *See* obviousness and compellingness
- compositionality, 208
- concepts. *See* apprehension of universals/ concepts/senses; conceptual realism; conceptualism; neo-Platonic account of abstract objects; possession conditions
- conceptual realism, 44
- conceptualism
- about justification vs. about truth, xxiii
 - about logic, xxiii–xxv, 77–122. *See also* constitution, justification by; indispensability: of logic; obviousness and compellingness; success, empirical
 - about universals, 243, 245, 250, 252–254, 257
 - See also* conceptual realism
- conditionals
- asserting, and deduction, 159–162, 168–170, 222–223
 - associated with dispositions, 188–189
 - and expressivism, 215–216
 - indicative vs. counterfactual/subjunctive, 147, 162
 - in if-then-ism, 40
 - modus ponens* as constitutive of understanding, 88–89
 - “normative,” 216
 - “sound,” 140n
 - validating, 166–167
 - See also* knowledge: of “not-*p* or *q*”
- conjunction, 84–85, 89–93, [93–95]n15, 96n, 98, 108, 143
- consciousness
- and naturalism about apprehension, xxxii, 237–240
 - of inferential steps, 112
 - phenomenal, 239–240
- consequence, logical. *See* logical consequence
- consequence relations, xxvi, 129, 133
- deductive capacity with respect to. *See* deductive capacity
 - first-order, 130–131, 198–199. *See also* first-order logic
 - latching onto, 130–131, 137
 - logic as comprising general laws of, 133–136
 - multiple-conclusion, 147–154
 - nature of, xxvi, 136–139
 - Philonian/material, 140, 144–146, 149, 153–154
 - relata* of, xxv–xxvi, 126–127
 - representational vs. interpretational notions of, 138
 - second-order, 131. *See also* second-order quantification/logic
 - Tarski's axioms for, 138–139, 144
 - See also* logical consequence; modality: and consequence
- consistency/inconsistency
- of commands, 221–222
 - formal, and coherence, 28
 - Frege and Hilbert on, 7–9
 - of generality absolutism, xxix, 202–203
 - and hyperstates, 218–219
 - and identifying structures, 32
 - of normative statements and judgments, 219–228
 - of set theory and second-order logic, 47
 - and “winning,” 226–228
- constitution, justification by, xxiii–xxiv, 78, 82, 91–103, 119–120
- constructible universe, 12, 59, 67
- constructive mathematics, 5–6

- contact/no-contact theory of knowledge acquisition, xiv, xix–xxii, xxxiii, 107.
See also causal accounts; causal contact theory of knowledge acquisition
- content
 and entitlement to transitions, 109
 and naturalistic accounts of apprehension, xxxii, 237–239
 of normative judgments, 217–220
 of normative statements, 208–211, 214–215
 outright *a priori*, 96
 of propositions, 127
 of pure descriptions and factual assessments, 259
- continuum
 categorical characterizations of, 2, 5
 Dedekind on the properties of, 6–7
 Dedekind's axiomatization of, 25
 determinateness/definiteness of, 56–57, 61–62
 with infinitesimals, 5
 intuitionistic theory of, 5
 structure of, 3n3, 5, 14–15, 21, 25, 29, 32
See also continuum hypothesis; real numbers
- continuum hypothesis (CH), 5, 11–18, 192
 and Borel Determinacy, 4–5, 53–54, 67
 compared with Kreisel's Λ_2 sentences, 54–55
 contrasted with Euclid's fifth postulate, 15–16, 48, 57, 59
 contrasted with the Axiom of Replacement, 15–16
 determinateness of, xxii, 4, 11, 13, 15, 17, 53–67
 generalized (GCH), 11, 53, 58, 62n
 as a genuine problem, 12, 53, 59
 independence of, 4, 12, 15–17, 48, 56–59, 66–67
 inherent vagueness of, 57
 new axioms for settling, 16, 48, 55
 and quasi-categoricity, xxii, 4, 11, 13, 15, 17, 58, 60–65
 Woodin's program for proving, 53–54, 59
- contraposition, 134, 144
- conventionalism, xvii, 240
- Copi, I.M., 129–131
- cornerstones/hinges, 78, 114, 116, 118, 121
- creativity at source, 254–255
- Csaba, F., 49n, 68
- cumulative hierarchy, 64
 and the Axiom of Foundation, 49–50
 characterization of, 10, 15, 38. *See also* quasi-categoricity
 containing all mathematical structures, 3, 39
 indefinite extensibility of, 45, 50–53, 67
 as intended model, 67
 notion/understanding of, 14, 55
 as a particular structure, xxii, xxx, 14, 38, 49–53
- cut law
 exhibited by all consequence relations, 139
 multiple-conclusion, 152–154
 Philonian, 153
- Darwin, C., 109–110
- Davidson, D., 111
- Davies, M., 173
- Dedekind, R., 2, 6–7, 11, 15–16, 20, 37, 41–42, 47, 56
 axioms for arithmetic, 24–25, 28, 36
- deduction, 151n17
 and asserting conditionals, 159–162, 168–170, 222–223
 and broadly logical validity, 157–158
 epistemological role of, xxvi, 141–145, 152, 167
 “free,” 197
 -like inference and deductive logic, 86
 rules for a universal quantifier, 193–198
 rules of, and truth-tables, 97
 and second-order logic, 34, 46–47
 from suppositions, 144–147, 156, 160
- deductive capacity, xxvi, 141–145, 160
- deep vs. superficial modality, 165–171, 173
- deflationism, xv–xvi, xviii, 79n3, 229. *See also* minimalism
- Dennett, D., 111
- Descartes, R., 43, 81, 106n22, 112, 117
- descriptive names, 162–166, 170–172

- Determination Theory, 96
discernibility. *See* indiscernibility/discernibility
domain-circularity. *See* single-domain problem
Dummett, M., 12, 68, 89, 93n14, 146n14, 187n12, 194n, 209n5, 216n, 220, 229, 231
on the notion of "object," 248–249
verificationism, xvi–xvii
- Edgington, D., 125n, 157, 160–164
enthymematic strategy, 129–131
emotivism, xiv, xxx, 214–215
empiricism, xx, xxv, 87, 132, 238–239
epistemic/non-epistemic modality, xxvii, 157, 159, 162, 169–170, 199–200
error-theory, xix
Etchemendy, J., 138
Euclidean geometry, 25
and Cartesian geometry 43
Euclid's axiomatisation of, 27, 30n
Hilbert's axiomatization of, 24, 27, 42
See also Euclidean plane; Euclid's fifth postulate; non-Euclidean geometry
Euclidean plane
categorical characterization of, 2–3, 7, 42
as a particular structure, 21–22, 29, 38, 42
points of, as objects, 42–44
See also Euclidean geometry
Euclid's fifth postulate
independence of, 12, 48, 53, 57
contrasted with CH, 15–16, 48, 57, 59
truth-value of, 25, 48
See also Euclidean geometry
Evans, G., 161–162, 165
Eynine, S., 106n22, 108n25
evolution, xxiv, 78, 81n, 109–11
exemplification/vacuity problem, 35–36, 41. *See also* Theorem 66
expressivism, xv, xxx–xxxii, 207
Blackburn's, 211–213
defeated, 225–232
Geach's Fregean problem for, 214–217
Gibbard's, 211, 213–214, 217–220
transformed into a factalist account, 232
and non-cognitivism, 208–211
and standards of normative discourse, 223–229
extension
Frege's notion of, 7, 249n
of *identical*, 177–179, 185, 190
externalism/internalism
and concept apprehension, 237–238
in epistemology of logic, 80–81, 85n
- factual assessment, 258–260
factalist semantics. *See* truth-conditional semantics
Feferman, S., 57, 62–63, 68
fictionalism, xvii, 236
in the neo-Platonic account of abstract objects, 242–243, 257–258
about normativity, 210n
about possibilities, 256
Field, H., xiv, xix, 56, 87n, 229n20
Fine, G., 245
Fine, K., 158n, 203n
first-order logic, xxvii, 179
consequence relation of, 130–131, 198–199
with/without identity, xxviii–xxix, 180n3, 181–182, 200
and second-order logic, 34, 46–47, 62, 186n11
and unrestricted quantification, 198–199
See also universal quantification/quantifier
forcing, 12, 59, 67
formal rigor, 13. *See also* informal rigor
forms, 30–31
individualized, 250–252
See also Plato's theory of Forms; universals
Foundation, Axiom of, 49–50
Fraenkel, A., 11, 16. *See also* Replacement, Axiom of; Zermelo-Fraenkel set theory
Frankfurt, H., 112–113
Frege, G., 18, 36, 42–43, 45, 47, 61, 84, 86, 93, [94–95]n, 209n5, 215
on arithmetic and numbers, 7, 241
Basic Law V, 7, 66
on concepts vs. objects, 243–245, 247–249

- Frege, G. (*cont.*)
 on consistency and existence, 7–9
 and reasoning from suppositions,
 146–147
See also Geach P.: Fregean problem;
 Julius Caesar problem
- Friedman, H., 5, 53, 198n24
- functionalism
 analytic, 81n
 and expressivism, 213
 Gibbard's, 217
- GCH. *See* generalized continuum hypothesis
- Geach, P., xxviii, 187, 191–192, 201
 Fregean problem, xxx–xxxi, 214–220
- general structures, 40
 vs. particular structures, xxi, 1–2,
 21–26, 34
- generality
 absolute, xxvii–xxx, 52–53, 192–203
 inconsistency/inarticulacy of absolutism
 about, 202–203
 relativism about, xxix, 52–53, 202–203
See also universal quantification/quantifier
- generalized continuum hypothesis
 (GCH), 11, 53, 58, 62n
- Gentzen, G., 135, 146n14
- geometry. *See* Euclidean geometry; non-Euclidean geometry
- Gibbard, A., xv, xvii, xxx–xxxi, 151n17,
 207, 211, 213–214, 217–220, 223–225,
 227, 232
- Glanzberg, M., 194n
- God, 254–257. *See also* objective intellect
- Gödel, K., xix–xx, 12, 13n, 16, 53, 55,
 59, 60–64, 88
 completeness theorem, 28, 46–47
 incompleteness theorems, 9, 12, 46,
 50, 60–62
See also von Neumann-Bernays-Gödel
 set theory
- Gödel sentence, 12, 54, 66
- Goldman, A., 79n3
- Goodman, N., 240
- grasp
 of absolute identity/generality, xxviii,
 192. *See also* open-ended commitment
 of abstract objects, xxxii, 235–237, 261
 cognitive, 87
 of conceptual truths, 96
 of inference patterns, 92
 of mathematical structures, 31
 pre-theoretical, of logical consequence,
 xxv, 126, 128–129, 133
 of validity, 88
See also apprehension of universals/concepts/senses
- Grover, D.L., xv
- Halbach, V., 68, 212n
- Hale, B., xx, 78, 89, 90n, 101, 104n, 216n
- Hare, R., xiv
- Harman, G., 117
- Hellman, G., xvii, 21, 30, 35, 40–41, 46
- Henkin, L., 186n11, 198n24
 completeness theorem, 46
- Hilbert, D., xv, 2, 7–10, 13n, 16, 18, 20,
 24, 27, 42
- hinges. *See* cornerstones/hinges
- Horgan, T., 211n
- Horwich, P., xv, 89, 103n, 229n20
- Humberstone, L., 173
- Hume, D., 237–238
- Hume's Principle, 241
- Hurley, S., 216n
- hyperinaccessible cardinals, 51–52
- hyperplans/hyperstates, xxxi, 218–220,
 225, 227, 232. *See also* plans
- identity
 absolute, xxvii–xxix, 177–193, 200–203
 between contents of utterances, xxv,
 126
 inarticulacy of absolutism about, 203
 logical vs. metaphysical necessity of, 158
 of mathematical objects and automorphisms, 42
 relativism about, 187
See also first-order logic: with/without
 identity; indiscernibility/discernibility; Leibniz's Law
- if-then-ism, 39–41
- in re* theory
 of structures, xviii–xix, xxi, 26, 30–31,
 33–37

- in re* theory (*cont.*)
 of universals, xxxiii, 33, 243, 251–253, 261
- inaccessible cardinals, 29, 50–51, 64
 hyperinaccessibles, 51–52
 strongly, 51–52
 Tarski's Axiom of, 51
- inarticulacy, xxix, 202–203
- inconsistency. *See* consistency/inconsistency
- indefinite extensibility, 45, 50–53, 67
- indiscernibility/discernibility, xxvii–xxviii, 183–186
- indispensability
 of logic, xxiv, 78–79, 82–83, 87, 90, 110, 112–119, 121
 of possible ω -sequences, 40
 Quine–Putnam argument, 79n3, 82, 87, 115
- individualized forms, 250–252
- infinity, axioms of, 55, 64
 Shapiro's, 27–28
 Zermelo's, 15, 37
- informal rigor, xxi–xxii, 13–15, 18, 29, 32, 34, 39, 41
- insight. *See* intuition/insight
- instances, 250–252
 grasping concepts and causal contact with, xxxii, 239, 241
 inferring universals from, 241–242
 universals as self-predicating, 244, 247
See also uninstantiated universals
- instantiation. *See* instances; self-instantiation/self-predication; uninstantiated universals
- intellect, xxxii–xxxiii, 235–237, 250, 261
 active, xviii, 250, 252
 objective, xxxiii, 242–243, 257–258, 261
- introduction/elimination rules
 for conjunction, 84, 89, 91–93, [93–95]n15, 143
 for implication, 89. *See also* *modus ponens*
 for the universal quantifier, xxviii, 193–198
- intuition/insight
a priori, xxxi
 logical, 80–82, 88–89, 98–99, 111, 120, 122
 mathematical, 64, 80–81, 87, 112. *See also* intuitionistic mathematics
 of natural numbers, 56
 rational, xx, 82, 88, 107
 of sets, 57, 64
See also obviousness and compellingness
- intuitionistic logic, 126, 135, 152, 231
- intuitionistic mathematics, xvii, 5–6, 56
- Isaacson, D., xi, xx–xxii, xxix, 2, 5n, 44, 50
- Jankins, C., 83n
- Jeshion, R., 106n22
- Julius Caesar problem, 8, 18–19
- justification
a posteriori/empirical, 78–79, 82–83, 99, 117–118, 121–122
 conceptualism about, vs. about truth, xxiv
 conceptualist, of logic. xiii–xxv, 77–122.
See also constitution, justification by;
 indispensability: of logic; obviousness and compellingness; success, empirical
 by constitution. *See* constitution, justification by
 Frege's, of *modus ponens*, 146–147
 of multiple-conclusion logic, 150
 not necessarily extended by deduction, 142–143
 standards of, for normative statements, 220–231
- Kalderon, M., 210n
- Kalmár, L., 12–13, 58
- Kant, I., 34, 51, 81, 102, 107, 110
- Kaplan, D., 178n2
- Katz, J., xx, 2
- Kitcher, P., xviii
- knowledge
a priori, ix, xxxi, 89, 92, 107, 110, 156–157
 causal contact theory of. *See* causal contact theory of knowledge acquisition
 common, sustaining an expression's use, 163–164, 171–172

- knowledge (*cont.*)
 contact/no-contact theory of. *See* contact/no-contact theory of knowledge
 acquisition
 by deduction, xxvi, 141–145, 152, 167
 of the continuum, 17–18, 56
 of Euclidean geometry, 43
 of logic, xxiii–xxv, 77–122, 146–147
 of logically necessary propositions, 156–157
 of mathematical objects, 1, 34
 of mathematical structures, xxii, 29, 36–37
 of mathematics, x, xii–xiii, xxxii, 10, 77
 of “not- p or q ,” 140, 144–145
 of numbers, 236, 241
 of particulars, 236, 241
 of set theory vs. of arithmetic, 66–67
 and supposing, 147
See also Benacerraf’s dilemma; skepticism/anti-skepticism
 Kreisel, G., xxi, 3n3, 12–16, 18, 20, 32, 45, 57–58, 61–64, 69
 incompleteness theorem, 54
 Kreisel Δ , sentences, 54–55
 Kripke, S., 157–158, 160–161, 187, 240
- L.* *See* constructible universe
 large cardinals, 53. *See also* inaccessible cardinals; Mahlo cardinals; Woodin cardinals
 Law V, 7, 66
 Leibniz, G.W., 88
 Leibniz’s Law, 178–179, 185–189
 Lemmon, E.J., 125–128
 Lenman, J., 222n, 229n20
 Lewis, D., xx, xxxv, 188nn, 204, 253n2
 Linnebo, Ø., 27n10, 37, 68, 203n28
 logic
 anti-realism about, 82, 86–87, 100–101, 120
 Benacerraf’s dilemma for. *See* Benacerraf’s dilemma: for logic
 epistemological role of. *See under* deduction
 first-order. *See* first-order logic
 free, 195–196
 “in the world,” 79, 86, 100, 120
 inclusive, 196
 indispensability of. *See* indispensability: of logic
 intuitionistic, 126, 135, 152, 231
 knowledge of, xxiii–xxv, 77–122, 146–147
 modal, 178, 213, 256
 multiple-conclusion, 147–154
 the nature of, xxv–xxvii, 125–173
 objectivity of. *See* objectivity: of logic
 realism about. *See* realism: about logic
 second-order. *See* second-order quantification/logic
See also logical modality; Platonism: about logical entities; reliability; suppositions
 logical consequence, xxvii, 154–173
 pre-theoretical grasp of, xxv, 126, 128–129, 133
 Tarski’s theory of, xxvi, 138, 199–200
 as uniquely favored, xxv, 128–133
See also consequence relations
 logical laws, 127
 conceptualist justifiers of, xxiii–xxiv, 114
 epistemological role of learning, xxvi, 143–144. *See also* deduction: epistemological role of
 as laws of a pre-theoretically grasped relationship, 126
 and rules of sequent calculi, 135–136
 suppositions contrary to, 156
 truth conditions of, xxvii
 as truths about all consequence relations, xxvi, 134
 logical modality, xxvii, 154–162, 165–170, 172–173
 logical truth, 178
 Benacerraf’s dilemma about, xi
 as byproducts of inference rules, 146
 conceptualism about, xxiii, xxv
 knowledge of, 147. *See also* knowledge: of logic
 realism about, xxiii, xxv
 referentialism/non-referentialism about, xxiii, xxv, xxvii
 Tarski’s account of, 196–197
 and unrestricted quantification, 196–199
 logicism/anti-logicism, 77, 86, 111

- logos*, being the same in, 251
 Löwenheim-Skolem Theorem, 10, 47, 59, 180n3
- Mackie, J.L., xxix
 Maddy, P., 60n, 61, 68
 Mahlo cardinals, 55
 material consequence. *See* Philonian consequence
 Mates, B., 125, 127–128
 mathematical intuition, 64, 80–81, 87, 112. *See also* intuitionistic mathematics
 mathematical objects, x–xiii, xx–xxii, 1–2, 21, 26
 axioms as conditions on, 15
 neo-Platonic account of, xxxiii
 as positions in structures, 20, 33, 41–42
 unreality of, 2, 9, 31–34, 37–38, 41–44
 See also numbers
 McFetridge, I., xxvii, 155–156, 158–162, 167–170, 172–173
 McGee, V., xxx, 190n, 194n
 metaphysical modality, 157–158, 161–162, 165–166, 170, 172–173
 metaphysics
 of concepts and senses, xxiii, 120
 and deflationist semantics, xv
 expressivism and naturalistic, 209
 of mathematics, xxii, 1–2, 44
 status of normative truth, 231
 See also metaphysical modality
 metasemantic account, 94–96, 107–108
more geometrico. *See* axiomatic method
 Mill, J.S., xviii, 142n, 155
 minimalism, xv, 212–213, 229
 Mišćević, N., xi, xxiii–xxv, 79n3
 modal logic, 178, 213, 256
 modal structuralism, xvii, xxi, 21, 26, 30, 41
 modality
 and consequence, xxvi–xxvii, 137–138, 140–147, 154, 156, 167, 199–200
 deep vs. superficial, 165–171, 173
 epistemic/non-epistemic, xxvii, 157, 159, 162, 169–170, 199–200
 epistemology of, 107–108
 logical, xxvii, 154–162, 165–170, 172–173
 metaphysical. *See* metaphysical modality
 and compellingness, 103
 physical, 137, 145, 165–166
 and theories of universals, 243, 252–257
 See also Combinatorial Theory of Possibility; modal structuralism; possibilities
 model theory, xxi, 9, 17, 39, 194, 199. *See also* model-theoretic notion of structure
 model-theoretic notion of structure, xxi, 26–27, 30–31, 35, 38–39
modus ponens, 84n, 88–89, 112, 132, 136, 146–147, 139n10, 216
 Moore, G., 11, 61, 63
 Moore, G.E., 128, 149n, 156
 Mostowski, A., 12–13, 47–48, 58–59
 multiple-conclusion cut, 152–154
 multiple-conclusion logic, 147–154
- Nagel, T., 111
 Nash, J.F., 112
 natural constitution claim, 214
 natural numbers, 9, 31, 37, 67
 categorical characterization of, 2, 32, 36–38, 47
 Dedekind and Frege on, 7
 Dedekind's axioms for, 24–25, 28, 36, 42, 56
 determinateness of problems about, 55–56
 knowledge of, by description, 241
 non-definable properties of, 16
 the notion of, vs. the notion of sets, 66
 as particular objects, 1–2, 7, 19, 37, 43
 quantification over, in intuitionistic mathematics, 55–56
 set-theoretic interpretations of, 3n3, 19, 42
 structure of, xxi, 1–2, 15, 21–25, 28–29, 31–32, 36–42, 47
 See also arithmetic; numbers
 naturalistic accounts, 79n3
 and Benacerraf's dilemma, xviii, xxxii, 81n, 235–237
 of concept apprehension, xxxiii, 237–242, 261
 and conceptualism, 89, 106, 108
 and expressivism, 209, 214, 231
 of logical abilities, 111

- NBG. *See* von Neumann-Bernays-Gödel set theory
- necessity. *See* modality
- negation, 108–109, 134–135, 138, 215
and allowing a hyperstate, 219
and “winning,” 227–228
See also propositions: negative, as assessments
- neo-Platonic account of abstract objects, xxxiii, 242–243, 257–258, 261
- non-circularity constraint, 208
- non-Euclidean geometry, 25, 57, 59
- non-standard models
of arithmetic, 28, 38, 67, 186n11
of second-order languages, 186n11
of set theory, 14, 17, 67
- normative statements, xxx–xxxii, 207–217, 220, 222n, 223–227, 230–232
- normative judgments, 207–208, 210–212, 217–220, 223–224
- Nous*. *See* objective intellect
- Novák, Zs., 68, 77n, 79n3, 98, 115
- numbers, 87, 157
and Benacerraf’s dilemma, xiii, 235–237, 261
complex, 22, 32, 38, 42
natural. *See* natural numbers
neo-Platonic account of, 242
real. *See* continuum; real numbers
- objective intellect, xxxiii, 242–243, 257–258, 261
- objectivity, 33
of *a priori* truths, ix
of concepts and relations, 63
of the continuum problem, 56
of logic, xi, 77, 79, 82, 86–87, 99, 102, 106, 120–121
of mathematics, 1–2, 5, 44
and obviousness and compellingness, xxiv, 121
of possibility/impossibility, 99, 253, 260
of senses, 102, 120
See also realism
- obviousness and compellingness, xxiv, 77–80, 82–84, 90–92, [93–94]n15, 99–108, 111–113
- Ogden, C.K., 239
- open-ended commitment, xxviii–xxix, 189–191, 193–198
- PA. *See* Peano Arithmetic
- paradoxes, 40
and absolute identity, 191–192, 201–202
Burali-Forti, 27, 202
logical, and multiple-conclusion frameworks, 152
Russell’s, 7, 66–67, 201–202
set-theoretic, 51, 192
of skepticism, 117–118
- particular structures, xxi–xxiii, 1–4, 26–44, 56
abstract vs. concrete, 30–31
vs. general structures, xxi, 1–2, 21–26, 34
- Peacocke, C., xvii, xxiv, 78, 82–83, 85–86, 89–104, 106n23, 107–110, 113, 119, 171n, 182n, 208n3
- Peano Arithmetic (PA), 5, 9, 24, 48, 54, 67
- Philo of Megara, 140n
- Philonian consequence, 140, 144–146, 149, 153–154
- Philonian cut, 153
- plans, xxxi, 217–220, 225, 227n. *See also* hyperplans
- Plantinga, A., 111
- Plato, 31, 81, 88, 235
theory of Forms. *See* Plato’s theory of Forms
- Platonic world/realm, xi, 102, 241–242, 257
- Platonism, 101
and causal contact, xxxii, 240–241
and instances, 252
Frege’s, 249n
full-blooded, xx
about logical entities, xi, 77–78
about mathematics, x, xviii–xxi, 1–2, 20, 42, 56, 63, 235–236
“modern,” 244, 247
object vs. concept, 44
about senses, 102, 120
about universals, xxxiii, 251–252. *See also* Plato’s theory of Forms; realism: about universals

- Platonism (*cont.*)
 See also *ante rem* theory; neo-Platonic account of abstract objects
- Plato's theory of Forms, 31, 244–247
 Aristotle's and Frege's criticism of, 243–245, 247–249
 See also forms
- Poincaré, H., 9, 18, 31
- possession conditions, xxiv, 92, 94–96, 101
- possibilities, 137–138, 260
 realism about, xxvii, xxxiii, 243, 252–254, 260
 truth *in/with respect to*, 137–138, 165–167, 169–170, 172
 See also Combinatorial Theory of Possibility; modality
- Pritchard, D., 116n
- propositions
 Benacerraf's dilemma for, xi
 negative, as assessments, 258
 normative, 212–213
 as pairs of contexts and sentences, xxv–xxvi, 127
 Platonic ontology of, xi, 77
 See also cognitivism/non-cognitivism; truth-conditional semantics
- Pryor, J., 115
- pure descriptions, 258–259
- Putnam, H., xvi–xvii, 40, 63, 79n3, 82, 87, 115
- qualities, 100–101, 250, 259
- quasi-categoricity, xxii, 3–4, 34, 50, 53, 58
 and the continuum problem, xxii, 4, 11, 13, 17, 64–65
 Gödel on, 60–64
 Shepherdson's inner model theorem, 4, 12, 49, 53
 and unrestricted quantification, xxx
 Zermelo's theorem, xxii, 3–4, 10–11, 17, 47, 49
- quasi-realism, xv–xvii, xxx–xxxix, 211–213
- quietism, xvii, xxiii, 101, 120
- Quine, W.V.O., xv, xx, 12, 44, 79n3, 82, 86–87, 99, 111, 114–115, 119, 132, 178n1
 criticism of second-order logic, 43, 47–48
 identification of indiscernibles, xxviii, 182–185
 See also Quineianism
- Quineianism, 242, 257
- Quine-Putnam indispensability argument, 79n3, 82, 87, 115
- Ramsey, F., xv, 13n
- rational insight, xx, 82, 88, 107
- Rayo, A., 194n, 196n22, 198n24, 203n
- real numbers, 57, 130–131, 181–182
 in Cartesian geometry, 43
 indefinable, 45
 as mathematical objects, 44
 theory of, and second-order arithmetic, 46
 See also continuum
- realism, 79n3, 99, 101
 ante rem. See *ante rem* theory. See also Platonism; Plato's theory of Forms;
 about concepts, xxiii
 conceptual, 44
 its contrast with anti-realism, xv–xvi
 in re. See *in re* theory
 internal, xvi
 about logic, xxiii, xxv, 77, 82, 87, 90n, 100, 115
 about mathematics, x, xiii, xviii–xxiii, xxxii, 1, 44, 55, 115, 235–236
 moral, 211
 about normativity, xxxi, 231
 physical, 259
 about possibilities, xxxiii, 243, 252–254, 260
 semantic, 231
 about senses, 120
 about truth, xii, xv–xviii, 231
 about universals, xxxiii, 33, 243–244, 251–253, 261
 See also anti-realism; Platonism; quasi-realism
- reciprocal conception, 93–98, 107
- reductionism
 about concept apprehension, xxxii, 238–239, 241
 set-theoretic, 3n3, 36, 59
 about normative truths, 231
 See also natural constitution claim

- referentialism/non-referentialism
 and Benacerraf's dilemma, x–xiii, xv–xix
 about logic, xxv, xxvii
 about mathematics, xxii–xxiii,
 about normative discourse, xxxi–xxxii,
 214
 reflective equilibrium, xxiv, 79–81, 98,
 102–103, 120
 rejection, attitude of, 148–152
 relativism
 about identity, xxviii, 187, 188n15
 about generality, xxix, 52–53, 202–203
 relativity of set-theoretic concepts, 49,
 52–53, 59
 reliability, 104n
 and contact, xiv, xx
 of a deductive capacity, xxvi, 142, 144
 of detecting obviousness, xxiv
 of empirical evidence for logic, xxv
 of logical principles and habits, 77–81,
 90, 99, 111
 of perception 98
 and success, 99
 Replacement, Axiom of, 4, 11, 15–17, 45,
 50, 54–55
 representations, xiii, xv, xvii, xx, 97
 causal theory of conceptual, xxxii,
 239–240
 resemblance, 250–251
 Resnik, M., xx, 20
 revisionism/non-revisionism, xiii, xviii–
 xix, xxii, xxv, xxxi, xxxiii, 5
 Rey, G., 106n23
 Ripps, L., 86
 Robinson, H., xi, xxxii–xxxiii, 254
 Rosen, G., 44, 208n2, 209n4, 224n
 rule-circularity, 90n, 104n, 114, 120.
 See also circularity: in justifications of
 logical principles
 Rumfitt, I., xi, xxv–xvii, 68, 126n, 142n,
 171n, 173n
 Russell, B., xxvi, 40, 42–43, 68, 77, 80–
 81, 84n, 85, 140–141, 143–147
 Russell's paradox, 7, 66–67, 201–202

 second-order languages, 28–29, 35, 39,
 45, 186n11. *See also* arithmetic: second
 order; second-order quantification/logic
 second-order quantification/logic, 44–49
 and absolutely unrestricted quantifiers,
 201
 categorical characterizations and, 2–3,
 34, 62–65
 and the enthymematic strategy, 131
 and first-order logic, 34, 46–47, 62,
 186n11
 and interpreting the identity predicate,
 185–186
 validity in, 158
 weak, 3n1
 See also second-order arithmetic; sec-
 ond-order languages
 self-instantiation/self-predication, 212n,
 244–247
 semantics. *See* cognitivism/non-cognitiv-
 ism; emotivism; expressivism; truth-
 conditional semantics
 sequent calculi, 135–136, 146n14
 set theory. *See* Zermelo–Fraenkel set theo-
 ry, Zermelo set theory; von Neumann–
 Bernays–Gödel set theory. *See also*
 cumulative hierarchy; structuralism:
 about set theory
 Shapiro, S., xx, 20, 33–34, 39, 41–44, 46,
 68–69, 185n9, 186n11
 theory of structures, xxi, 26–30
 Axiom of Infinity, 27–28
 Shepherdson, J., 4, 12, 29, 49, 53
 Simonyi, A., 68, 77n, 79n3, 115
 single-domain problem, 97–98, 100, 120
 Sinnott-Armstrong, W., 224
 skepticism/anti-skepticism, 113–115,
 117–118
 and Benacerraf's dilemma, xiii, xix
 about identity, 187
 about obviousness and compellingness,
 83, 107, 112–113, 121
 Skolem, T., 10, 11, 49, 59. *See also*
 Löwenheim–Skolem theorem
 Smiley, T., 129, 130n, 132n5, 150, 152
 Smokrović, N., 98n
 Sosa, E., 80n6
 Sprigge, T., 259
 Stevenson, C.L., xiv
 Stich, S., 111

- Strawson, P., xv, 132n4
 Stroud, B., 88n
 structuralism, ix–xi, xx–xxi, 6–7, 9–10, 18, 20–21
 ante rem, xviii–xix, xxi–xxii, 31–34, 37, 42
 and Benacerraf's dilemma, x–xi, xvii–xxi,
 in re, xviii–xix, xxi, 26, 30–31, 33–37
 and mathematical objects, x–xi, xx–xxii, 1–2, 9, 20–21, 31–34, 37–38, 41–44
 modal, xvii, xxi, 21, 26, 30, 41
 about set theory, xxii, xxix–xxx, 3–5, 29–30, 49–53
 Shapiro's, xx–xxi, 20–21, 26–30, 33–34, 41–44
 See also exemplification/vacuity problem; if-thenism
 structures. *See* general structures; model-theoretic notion of structure; particular structures
 substantivism about truth, xii, xv–xvi, xviii, xxxi
 success, empirical, xxiv, 79, 83, 87, 99, 102, 116–118, 121
 supervenience, 213, 223
 suppositions
 and broadly logical consequence, xxvi–xxvii, 154–156, 171–172, 167–168
 counterfactual, 160, 171
 and propositions, 127
 reasoning from, xxvi–xxvii, 144–147, 154–156, 160, 222–223

 tacit premises, 129–131
 Tarski, A., xiii, 12
 theory of consequence, xxvi, 138–139, 144, 196–200
 model-theoretic notion of structure, 26, 35
 Axiom of Inaccessible Cardinals, 51
 Tarski's schema, xv, xxxi, 228n
 Tennant, N., 149, 231n
 Theorem 66, 37, 41
 Timmons, M., 211n
 tropes, 250–251
 truth
 a priori, ix, xvii. *See also* content: outright *a priori*
 arithmetical, 50
 deflationism about, xv–xvi, 79n3, 229
 in/with respect to a possibility, 137–138, 165–167, 169–170, 172
 logical. *See* logical truths
 mathematical, x, xii, xxi–xxiii, xxxii, 1
 minimalism about, xv, 212–213, 229
 normative, xxxi, 231
 realism/anti-realism about, xii, xv–xviii, 101, 231
 substantivism about, xii, xv–xvi, xviii, xxxi
 truth-conditional semantics, xxx–xxxii, 207, 209–211, 213–214, 217, 229, 231–232
 truthmakers, 257, 259–260

 undecidable statements, 60–61, 66. *See also* Gödel sentence; Kreisel Δ_2 sentences
 uninstantiated universals, xxxiii, 252–256
 unique characterization
 of identity, xxviii, 190, 192
 of structures. *See* categorical characterizations
 of the universal quantifier, xxviii, 193–195
 universal quantification/quantifier
 constant/variable account of unrestricted, xxix–xxx, 197–200
 open-ended rules for, xxviii–xxix, 193–198
 over natural numbers in intuitionistic mathematics, 55–56
 over structures, 35–36
 unique characterization of, xxviii, 193–195
 unrestricted, xxviii, 52, 197–201
 See also generality; second-order quantification/logic
 universals
 acquaintance with, 85n, 241
 and instances, 252
 ante rem theory of. *See ante rem* theory: of universals
 apprehension of, xxxii–xxxiii, 235–242, 261
 conceptualism about. *See* conceptualism: about universals

- universals (*cont.*)
 - empirical, xxxii, 241
 - in re* theory of. *See in re* theory: of universals
 - neo-Platonic account of, xxxiii, 242–261
 - objectification of, 244–249
 - ontological status of, xxxiii
 - Platonism about. *See* Platonism: about universals
 - realism about. *See* realism: about universals
 - resemblance theory of, 250–251
 - uninstantiated, xxxiii, 252–253, 256
- V*. *See* cumulative hierarchy
- $V=L$ axiom, 59
- vacuity. *See* exemplification/vacuity problem
- von Neumann-Bernays-Gödel set theory (NBG), 4, 12, 49
- Walker, R.C.S., 240
- Wedgwood, R., xi, xxx–xxxi, 207n, 213n10, 229n20
- Wiggins, D., 184n, 187n12, 229
- Williams, B., 112
- Williamson, T., xi, xxiv, xxvii–xxx, 68, 79n4, 83n, 96, 190n, 191n, 196nn, 198n24, 201n
- “winning” statements/sentences, 226–229, 232
- Wittgenstein, L., 13n, 100, 114–115, 240
- Woodin, H., 69
 - program, 53–54, 59
- Woodin cardinals, 54
- Wright, C., xx, xxxi, 27n11, 38, 55, 68, 78, 82–83, 86, 88–90, 101–102, 112–119, 125n, 220
- Zermelo, E., 3–4, 10–12, 14–17, 19–20, 29, 30n, 37, 47, 49–55, 58, 60–64.
 - See also* Zermelo-Fraenkel set theory; Zermelo set theory
- Zermelo-Fraenkel set theory
 - first-order, 49–50, 65
 - CH’s independence from, 12–13, 16, 48, 58–59
 - second-order
 - and CH, 4–5, 13, 15–17, 48, 57–58, 60–65
 - coherence of, 29
 - incompleteness of, 58
 - quasi-categoricity of. *See* quasi-categoricity
 - with urelements, xxx
- Zermelo set theory, 14–16, 30n, 53–54.
 - See also* Zermelo-Fraenkel set theory
- ZF, ZFC. *See* Zermelo-Fraenkel set theory: first-order
- ZF², ZFC². *See* Zermelo-Fraenkel set theory: second-order
- Zvolenszky, Zs., 77n

WHAT MAKES IT THE CASE THAT A CERTAIN BELIEF IS TRUE? DO WE HAVE AN APPROPRIATE CONCEPTION OF THE TRUTH CONDITIONS OF OUR BELIEFS? CAN WE EXPLAIN HOW WE DISCOVER THE OBTAINING OF THESE CONDITIONS? AND, IF THE ANSWERS TO THESE QUESTIONS ARE POSITIVE, ARE THESE THEORIES OF TRUTH AND KNOWLEDGE ACQUISITION COMPATIBLE WITH EACH OTHER AND OUR SCIENTIFIC UNDERSTANDING OF HUMANS IN THE WORLD? IN THE LAST FEW DECADES, THERE HAS BEEN A GROWING SENSE IN PHILOSOPHY OF THE DIFFICULTIES ONE HAS TO FACE WHEN TRYING TO ENSURE THIS COMPATIBILITY IN THE SEMANTICS AND EPISTEMOLOGY OF VARIOUS DISCOURSES, ESPECIALLY THOSE IN WHICH WE ARE SUPPOSED TO ACQUIRE KNOWLEDGE ABOUT CAUSALLY INERT SUBJECT MATTERS. THE SIX PAPERS COLLECTED IN THIS VOLUME ADDRESS THE ABOVE CONCERNS IN THE CONTEXT OF MATHEMATICAL, LOGICAL AND ETHICAL BELIEFS, AS WELL AS IN THE CASE OF UNIVERSAL GENERALISATIONS AND BELIEFS INVOLVING IDEAS OF UNIVERSALS IN GENERAL.

Central European University Press
Budapest – New York
Sales and information: ceupress@ceu.hu
Website: <http://www.ceupress.com>

